(12)

Technical Report 573

AD-A133 135

OPTIMIZING OPERATOR PERFORMANCE ON ADVANCED TRAINING SIMULATORS:

PRELIMINARY DEVELOPMENT OF A PERFORMANCE ASSESSMENT AND MODELING CAPABILITY

John K. Hawley
Applied Science Associates, Inc.

Charles W. Howard
Army Research Institute

A. John Martellaro
Applied Science Associates, Inc.

ARI FIELD UNIT AT FORT BLISS, TEXAS

U. S. Army

Research Institute for the Behavioral and Social Sciences

February 1982

Approved for public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

Research accomplished under contract
to the Department of the Army

Applied Science Associates, Inc.

NOTICES

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report 573	2. GOVT ACCESSION NO.: AD-A133 135
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): OPTIMIZING OPERATOR PERFORMANCE ON ADVANCED TRAINING SIMULATORS: PRELIMINARY DEVELOPMENT OF A PERFORMANCE ASSESSMENT AND MODELING CAPABILITY	5. TYPE OF REPORT & PERIOD COVERED: Technical Report May 1980 - February 1982
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): John K. Hawley, Charles W. Howard and A. John Martellaro	8. CONTRACT OR GRANT NUMBER(s): MDA903-80-C-0506
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Applied Science Associates, Inc. Box 158 Valencia, PA 16059	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q263744A795
11. CONTROLLING OFFICE NAME AND ADDRESS: US Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Ave., Alexandria, VA 22333	12. REPORT DATE: February 1982
	13. NUMBER OF PAGES: 78
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): ARI Field Unit P.O. Box 6057 Ft. Bliss, TX 79916	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

--

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Human-machine system; PATRIOT Air Defense Missile System; Performance Optimization; Performance Measures; Model Validation; Operator Training; Situational Difficulty Index; Simulation Model; Operator Selection; Performance Evaluation; Human-machine Integration; Model Parameterization

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents results from the first phase of a project concerned with human-machine integration in computer-aided systems, with a specific focus on the PATRIOT Air Defense missile system. The objective of the work is the development of a performance optimization model that will relate human-operator performance to aspects of operator training and selection, human-machine integration, and system deployment.

To this end, the first set of project activities concern the quantification of operator performance in the PATRIOT system. A hierarchy of performance —

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

Con't

measures characterizing operator performance at the system, mission, and individual task levels was developed and implemented on a PATRIOT environmental, full-task simulator. Procedures for adjusting raw operator scores to ~~reflect environmental~~ scenario difficulty were also developed.

The second set of project activities involve the development of a simulation model of a PATRIOT Engagement Control Station console operator. This simulation model is to be used as a partial surrogate for experimentation with actual console operators in the construction of an operator performance optimization model. The logical basis for the operator model is described and procedures for parameterizing and validating the model are presented. Results from a preliminary implementation of the model are also discussed. Finally, future directions for the performance measurement and modeling activities are noted.

Technical Report 573

OPTIMIZING OPERATOR PERFORMANCE ON ADVANCED TRAINING SIMULATORS:

PRELIMINARY DEVELOPMENT OF A PERFORMANCE ASSESSMENT AND MODELING CAPABILITY

John K. Hawley
Applied Science Associates, Inc.

Charles W. Howard
Army Research Institute

A. John Martellaro
Applied Science Associates, Inc.

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

Submitted by:
Michael H. Strub, Chief
ARI FIELD UNIT AT FORT BLISS, TEXAS

Approved by:
Stanley M. Halpin, Acting Director
SYSTEMS RESEARCH LABORATORY

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

Army Project Number
2Q263744A795

Training Simulation

February 1982

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The US Army Research Institute for the Behavioral and Social Sciences (ARI) has, as part of its mission, provided R&D support to the US Army Air Defense School (USAADS) and the PATRIOT project Management Office. Part of the ARI effort has been to research human-machine integration in computer-aided systems. ARI initiated a research program on PATRIOT console operators performance analysis in 1978. The research objective was to develop performance optimization criteria that would relate operator performance to aspects of operator training and selection, human machine integration, and system deployment. The result of this effort was the identification of a hierarchy of performance measures quantifying system, mission, and individual task performance. The measures were subsequently implemented on a PATRIOT environmental, full-task simulator. Procedures for normalizing raw operator performance scores to adjust for the scenario level of complexity or threat load are presented. This research is in response to Army project 2Q263744A795 and special needs of the Directorates of Combat and Training Developments, USAADS and the PATRIOT Project Management Office.

Edgar M Johnson

EDGAR M. JOHNSON
Technical Director

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution to the project of Mr. John Armendariz, Mr. Syl Carrasco, CPT William Veitch, CPT Dave Smull, CW4 Robert McFall, and Specialist Dan Arvieux of USAADS DCD. Without their support and assistance, the programming work carried out in this project would have been a much more difficult undertaking.

Special recognition is also due for Mr. Donald Harris of the PATRIOT Deployment Support Office at Ft. Bliss. His subject matter expertise and advice were an invaluable asset to the project staff.

Optimizing Operator Performance on Advanced Training Simulators:
Preliminary Development of a Performance Assessment and Modeling Capability

BRIEF

Requirement:

The modern weapon systems of the twentieth century are challenging the limits of the human in the area of man-machine integration. To minimize degradation of system performance, it is necessary to enhance human functioning. This report describes the performance metrics developed in order to evaluate human functioning, system performance, and the scenario or situational difficulty for the PATRIOT Air Defense Weapon System.

Procedure:

The design and development of measures of effectiveness or performance metrics for the mission, system, console operator tasks, and environmental or scenario factors were required. Following the validation of the measures of effectiveness, each measure was implemented on a PATRIOT environmental, full-task simulator located at USAADS, Ft Bliss, Texas. Procedures were developed to normalize operator raw scores to reflect environmental scenario difficulty. The construction of an operator performance optimization model was then feasible, given the availability of the PATRIOT simulator, measures of effectiveness, and indices of scenario difficulty.

Findings:

Considerable ground work in the areas of operator performance assessment and modeling was laid. In the area of performance evaluation, probably the most important contributions are: (1) a clarification of Meister's framework for human-machine performance evaluation, and (2) an application of this framework using the PATRIOT Air Defense Missile System as an exemplary. In this exemplary application, it was demonstrated that operator performance can be quantified at a variety of levels and that the resulting data are reasonable.

A second major contribution of the current project concerns the treatment of situational difficulty as a modifier of operator performance. Previous efforts have recognized the necessity of adjusting raw operator performance indices to reflect situational difficulty, but satisfactory methods for treating the problem were not forthcoming. The treatment of situational difficulty as described herein is not definitive, but the approach holds promise for the future.

Utilization of Findings:

This report presents results from the first year of a research effort concerned with the general topic of human-machine integration in automated systems. The specific focus of the effort is on the PATRIOT Air Defense Missile System, but the methodology holds promise for application in other

computer-aided human-machine systems, such as air traffic control, nuclear power plants, or anti-submarine warfare. As noted several times during the report, the primary thrust of the effort is the development of a vehicle for enhancing overall system performance through a systematic consideration of performance shaping factors such as: (1) operator selection, (2) operator training, (3) human-machine integration, and (4) system employment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

In recent years, the United States has been faced with an unprecedented build-up of Soviet military power. Studies of the comparative military strengths of the United States and its allies (i.e., NATO) and the Warsaw Pact countries indicate that in a prospective engagement NATO forces are likely to be heavily outnumbered, often by a ratio of five to one or more. It is generally believed that the balance of forces in terms of numbers of tactical aircraft is particularly overwhelming in favor of the Soviet bloc.

In response to this trend, the United States has developed and is about to field a new generation of sophisticated weapons systems designed to meet and defeat the Soviet threat. Not ignoring the growing Soviet air threat, the U.S. Army has been in the forefront of this process with the development of a series of new air defense systems, such as Stinger, Roland, DIVAD, and PATRIOT. Several of these new systems represent a distinct break with the past. Many functions previously carried out by human operators and maintenance personnel now are automated or computer-aided. On the operations side, activities such as target identification, target tracking, weapons assignment, and target engagement are performed in an automated environment where computers have either displaced or greatly assist humans. The coming of these computer-aided systems has necessitated a reappraisal of the human operator's role in an automated air defense human-machine environment. Specifically of interest in this reappraisal are topics relevant to four problem areas, listed as follows:

1. Operator selection
2. Operator training
3. Human-machine integration
4. System employment doctrine

The U.S. Army Research Institute (ARI) Field Unit at Ft. Bliss, Texas, with contractor support from Applied Science Associates, Inc. (ASA), has initiated a research program concerned with various means of enhancing total human-machine system performance through a rational consideration of the four topic areas noted above. This report presents results from the first year of a research program concerned with these issues, with a specific emphasis on the PATRIOT Air Defense missile system. During the research program, five technical objectives, all relevant to the PATRIOT Engagement Control Station (ECS) console operator, were addressed. These technical objectives are listed as follows:

1. Development of a quantitative measure of effectiveness for overall console operator performance.

2. Determination of measures of effectiveness relative to specific console operator tasks.

3. Development of a procedure for the optimization modeling of PATRIOT console operator performance.

4. Validation of the procedure for modeling console operator performance.

5. Application of the optimization modeling procedure.

Objectives three, four, and five represent the crux of the effort directed at enhancing overall human-machine performance for the PATRIOT system. What is desired is a procedure for relating various aspects of operator selection, operator training, system integration, and system employment to total system performance. Before considering performance enhancement issues, however, it is first necessary to quantify operator performance in the human-machine system. These measures serve as the criteria for effective operator performance in the treatment of objectives three, four, and five.

Report Overview

The two primary topic areas noted in the previous paragraph--operator performance assessment and operator optimization modeling--serve to organize the presentation of material in the report. Section 2 addresses technical objectives one and two, i.e., the definition of a network of operator performance measures. In this section, a theoretical framework for console operator performance evaluation is presented and a network of quantitative performance measures is defined. The concomitant issue of situational difficulty as a moderator of operator performance is also examined. Finally, the results of an initial implementation of the performance assessment capability on an environmental, full-task simulator for the PATRIOT system are presented.

The subject of section 3 is the development of the operator optimization model. Section 3 begins with a general discussion of the operator performance prediction problem. Next, an approach to generating the data required to structure an appropriate performance prediction model is presented. This approach is based upon the development of a simulation model of a PATRIOT ECS console operator. Once developed and validated, the PATRIOT operator model is to be used as a partial surrogate for experimentation with actual console operators. Work carried out thus far in conceptualizing, developing, parameterizing, and validating the operator simulation model is also reviewed.

Finally, section 4 presents a summary overview of the report along with a discussion of successes, failures, and lessons learned. The last portion of section 4 is concerned with future directions for the performance assessment and modeling work initiated under the current effort.

Prior to presenting a detailed discussion of project activities, an overview of the structure and operation of the PATRIOT Air Defense missile system is provided in the next series of paragraphs. Since the PATRIOT system provides the context for the effort, it is desirable that the reader have at least a passing appreciation for the system's concept of operation.

The PATRIOT System

PATRIOT is an air defense missile system designed to combat the air threat of the late 1980s and 1990s. The system is intended to replace the aging Nike-Hercules system and to augment the HAWK system. Figure 1-1 depicts the structure of a typical PATRIOT battalion. As noted in Figure 1-1, a PATRIOT battalion consists of a headquarters command and coordination element and three firing batteries. Each firing battery consists of two firing platoons that each has the capability of directing up to eight launching stations. In PATRIOT, the battery-level element serves no direct tactical function, rather it serves as an administrative adjunct of the battalion [FM 44-15-1 (Test)].

During an air defense engagement, the launching stations are directed by the Engagement Control Stations (ECSs). The PATRIOT ECSs can be operated in either of two modes: automatic or semi-automatic. In automatic mode, the weapons control computer (WCC) in the ECS is able to direct most aspects of the air defense mission without human intervention. Since the engagement process is automated, firing doctrine is followed explicitly; threat evaluation and engagement decisions are performed in a theoretically optimum fashion.

The semi-automatic mode of operation introduces a human element into the PATRIOT engagement process. In semi-automatic mode, many of the functions performed by the ECS computer in automatic mode are carried out by a team of human console operators, each of whom is able individually to direct all ECS operations. The human operators are assisted by the WCC (if requested), but decisions concerning the sequence and timing of target engagements are made by the operators alone.

Given the response latencies, error probabilities, and information processing limitations of humans, it is expected that semi-automatic system performance will fall short of the level attained by the system in automatic mode ("Final Report Prototype," 1980). As attractive as the automatic mode of operation might be, however, concern over accidental engagement of friendly aircraft and a general distrust of automated systems have resulted in a decision that the PATRIOT system will operate in semi-automatic mode unless specific directives to the contrary are issued. Current doctrine specifies that human operators will participate in all decisions to launch missiles.

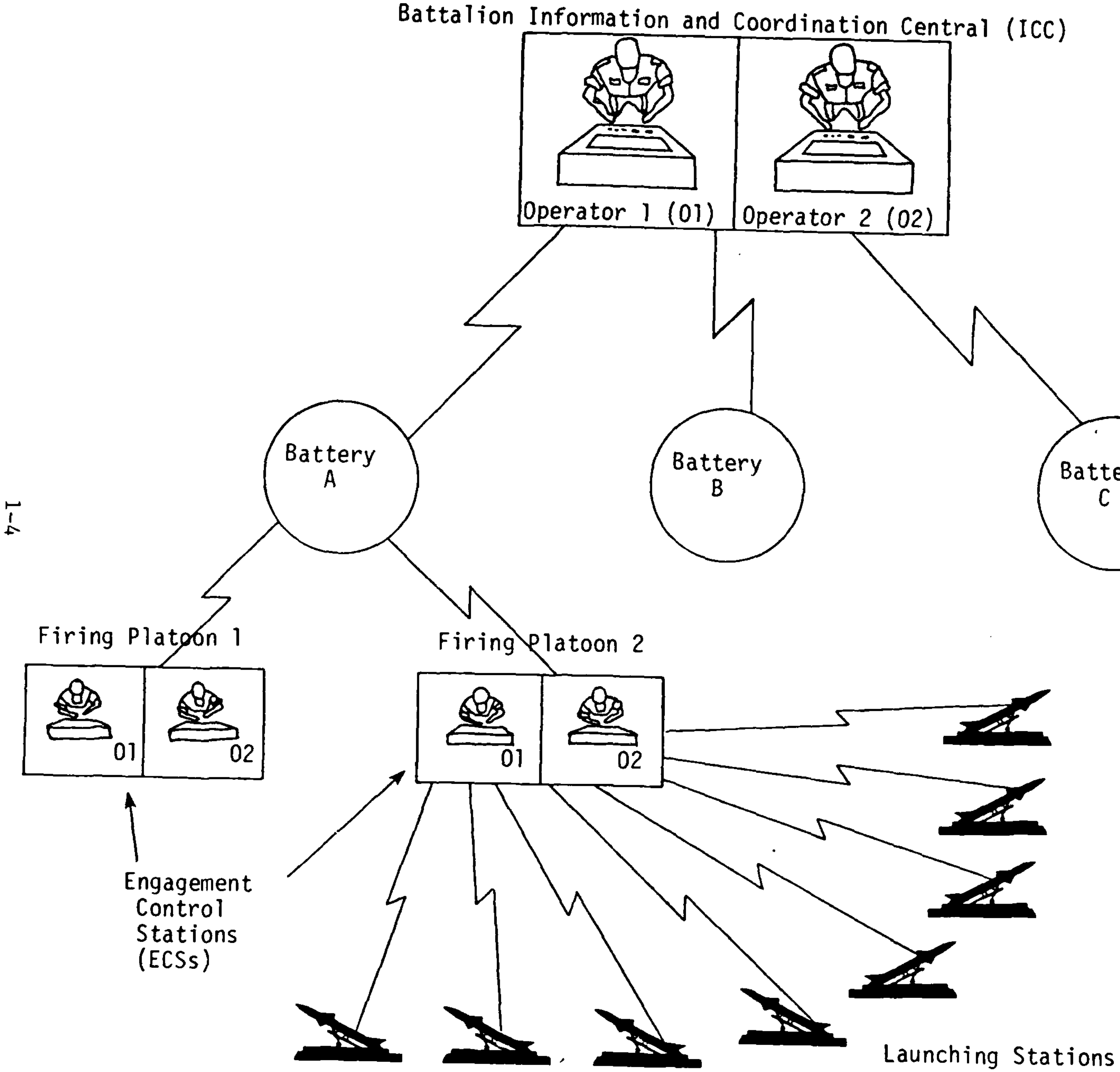

Figure 1-1. PATRIOT Concept of Operation

The activities of the six PATRIOT firing platoons are coordinated through a battalion-level Army Air Defense Command Post (AADCP). Located at the AADCP is the Information and Coordination Central (ICC). The ICC is similar in physical layout and function to the ECS, but it does not interface directly with radars or with launching stations. The primary functions of the ICC are: (1) correlation and management of track inputs from outside sources; (2) selection of firing platoons for engagement; and (3) management of external interfaces (e.g., to adjacent battalions or to higher headquarters). Like the ECS, ICC activities are monitored and controlled by two console operators, each of whom is able individually to direct all ICC operations. PATRIOT firing platoons are, however, also capable of functioning independent of the ICC.

The next section of the report addresses the first general topic area of the study: the development of a performance assessment scheme for PATRIOT ECS console operators.

2. OPERATOR PERFORMANCE ASSESSMENT

The first requirement in the current project concerns the development of criteria defining effective PATRIOT ECS console operator performance. This general requirement subsumes two separate technical objectives:

1. Establish a quantitative effectiveness measure describing overall ECS console operator performance.

2. Determine measures of effectiveness relative to specific console operator tasks.

As noted in the introductory section, these performance measures are to serve as the criterion context within which to construct the operator performance optimization model.

Following a systematic review of the literature relevant to operator performance measurement in command, control, and communications (C^3) systems (e.g., Sheldon & Zagorski, 1965; Howard, 1978; Howard, 1979; Jorgensen & Strub, 1979; Hopkin, 1980; Rouse, 1980), a decision was made to develop the PATRIOT console operator performance measures within a framework described in Meister (1976). To begin his discussion of performance assessment, Meister notes three characteristics of human-machine systems (PATRIOT most certainly qualifies as a human-machine system) that should influence performance measurement in such systems. The first of these characteristics is that human-machine systems are goal-oriented. This feature requires that, to appropriately assess human performance in a human-machine system, it is necessary to consider first the system's goals. Since the operator is a subsystem of the total human-machine system, the system's goals serve to define the operator's function. An operator is effective when his actions serve to meet the system's goals; he is ineffective when his actions do not serve the system's goals.

A second characteristic of human-machine systems relevant to human operator performance assessment is that such systems are typically hierarchically organized. That is, the total system is composed of subsystems, with the subsystems being, in turn, composed of subsystems at lower levels. The current project treats only ECS operators and does not consider operator interactions with other PATRIOT system components. To have utility in a more general context, however, the performance assessment scheme developed for a single-station ECS operator should recognize the hierarchical structure of the total PATRIOT human-machine system.

The third characteristic of human-machine systems that impacts upon performance assessment is the system's determinacy. With determinate systems, operator actions are prescribed through a structure imposed by the system. For example, the occurrence of stimulus X always prompts

operator response Y; other responses are inappropriate. Indeterminate systems, on the other hand, require the human operator to make choices among responses. Again using the above example, stimulus X may be followed by any of several legitimate responses, some of which may be more appropriate than others. The performance of the operator in the human-machine system cannot be adequately evaluated before the indeterminacies associated with each level of the system are noted and characterized.

Meister continues his discussion of human-machine performance assessment by noting three separate levels of performance criteria. These levels reflect conceptually different requirements for human machine performance assessment and are listed as follows:

1. System-(or subsystem) descriptive criteria: reliability, maintainability, acceptability, effectiveness (output), and efficiency.

2. Mission-descriptive criteria: output quantity and accuracy, reaction time, and queues and delays.

3. Personnel performance criteria: criteria associated with individual operator or crew behavior such as reaction time, response accuracy, response number, speed, variability, and so forth.

The three levels of performance criteria noted above can be thought of as defining a performance hierarchy. System-descriptive criteria represent the apex or top node in the hierarchy. These measures characterize the performance of the total system (or subsystem) with human operators in the control loop. Nested under the system-descriptive level are the various mission components. The mission-descriptive components, when integrated, serve to define total system performance. The lowest level of the performance hierarchy consists of personnel performance criteria. System-descriptive criteria and nested mission criteria serve to define the criteria that govern the operator's performance on individual tasks.

Figure 2-1 presents a schematic representation of Meister's performance measure hierarchy. In Figure 2-1, the arrows connecting the mission components to the system-descriptive level reflect the fact that mission performance directly relates to system performance. The links between personnel task performance and mission and system performance are, however, more tenuous in nature. Hence, the absence of a direct connection between the personnel performance level and the mission-/system-descriptive levels. It can be stated that acceptable performance of individual operator tasks is a necessary but not sufficient condition for acceptable performance at the mission- or system-descriptive levels.

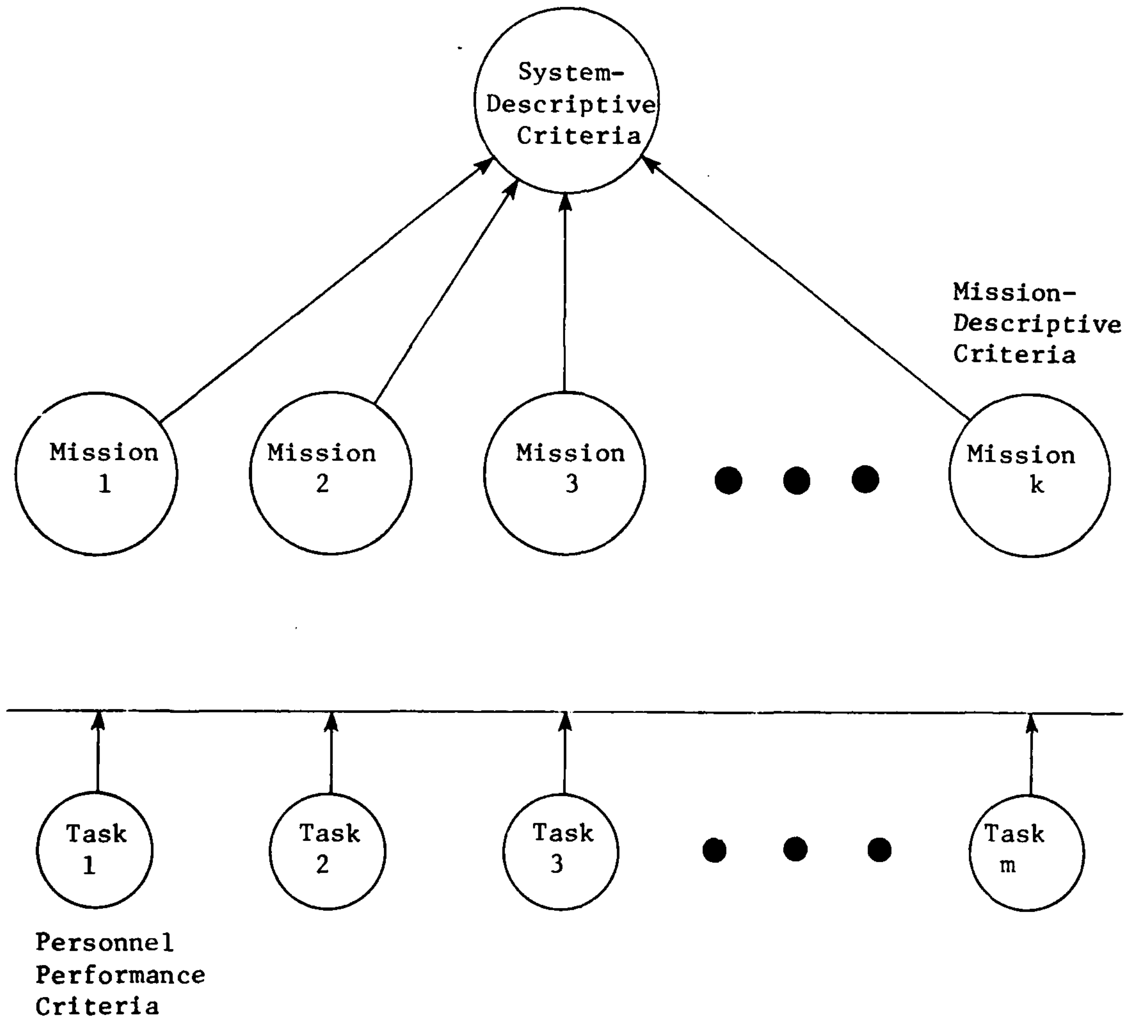

Figure 2-1. Schematic Representation of Meister's Hierarchy of Performance Measures

Development of Performance Measures

Following the conceptual framework outlined above, the first step in defining a network of performance measures for ECS console operators consists of defining the ECS subsystem's objectives and then characterizing associated determinacies/indeterminacies. Performance assessment issues associated with the fact that the ECS comprises only one level of a hierarchical human-machine system are not directly considered in the development of this initial series of operator performance measures. The next portion of the report describes the development of the performance measures characterizing the system- and mission-descriptive levels of the performance hierarchy.

System/Mission Performance Measures

In order to characterize operator performance at the system-descriptive level, it is necessary to consider first the overall objective of the PATRIOT system. The overall objective is then broken down into a series of sub-objectives that form the basis for the definition of mission-descriptive performance criteria.

A review of PATRIOT deployment doctrine indicates that the system, in essence, has two potential modes of application: as a point defense weapon and as an area defense, or attrition, weapon (FM 44-15). In the former mode, PATRIOT is employed to defend specific assets, such as an airfield or a command post. The system's objective in these instances is to prevent physical asset penetration by non-friendly aircraft.

As an area defense weapon, the PATRIOT system is charged with the management of a designated area. All non-friendly aircraft in this area of responsibility are "fair game", so to speak. The system's objective under this mode of application is to eliminate enemy aircraft in its defensive area. It is also possible for PATRIOT to be employed in a combination area defense-point defense mode, thus having two system-level objectives.

As part of a previous effort to define a performance evaluation scheme for the PATRIOT system in automatic mode (Hosch, Starner, and Howard, 1980), system performance was characterized as a function of three components, reflecting the system's mode of application: defense of assets, maintenance of defensive position, and effective missile utilization. Overall system, or summary, performance was defined as a weighted composite of the three components. After reviewing this earlier work, and given the decided upon framework for the development of performance measures (i.e., Meister), a decision was made to follow a similar approach in defining a network of <u>operator</u> performance measures for PATRIOT. Specifically, the three performance components listed previously were selected to define the mission-descriptive level of the performance hierarchy; the system performance

measure is then defined as a weighted composite of the three mission-level components. Conceptually, the measures are similar to those developed by Hosch, Starner, and Howard, but operationally they are defined to reflect human performance as opposed to human-machine system performance.

Following Hosch, Starner, and Howard (1980), the three performance components characterizing operator performance at the mission-descriptive level are defined as follows:

1. Maintenance of Defensive Position (DP). Timely processing of non-friendly (i.e., hostiles and unknowns eligible for engagement) aircraft as they enter the defensive envelope (i.e., enter the ECS station's area of responsibility and attain engageable status). DP may be thought of as effective management of the ECS station's airspace volume of responsibility.

2. Defense of Assets (DA). Protection of defended assets against physical penetration by hostile aircraft.

3. Resource Utilization (RU). Efficient use of defensive resources (i.e., missiles) in meeting mission requirements.

Having conceptually defined the system-/mission-descriptive level, the next step in the development of system/mission performance measures is operational definition and quantification. It is necessary, first, to define the performance constructs in terms of observables within the PATRIOT human-machine environment and then to develop a quantitative scoring rule for each. The three components defining the mission-descriptive level of the performance hierarchy are operationally defined and quantified in the following paragraphs.

Maintenance of Defensive Position. DP is a measure of the extent to which an operator engages eligible, non-friendly tracks in a timely fashion. Quantitatively, the measure is given as:

$$DP = \left[1 - \frac{\sum^{N_h} (TNE_i/TEE_i)}{N_h} \right] * 100. \qquad (2\text{-}1)$$

In (2-1), TNE_i (Time Not Engaged) is the elapsed time that non-friendly track i is eligible for engagement but not engaged. TNE_i is equal to the time track i is engaged (Et) or exits the To Be Engaged Queue (TBEQ) minus the time when track i became eligible for engagement (EEt).

TEE_i (Time Eligible for Engagement) is the elapsed time from when track $\underline{i}$ became eligible for engagement until either launch time or the time at which track $\underline{i}$ is no longer engageable (i.e., is no longer on the TBEQ). If track $\underline{i}$ is launched on, then $TEE_i = Et_i - EET_i + 1.5$, where 1.5 is a fixed minimum system delay time for launch. In this manner, the operator is not penalized for launch delay time due to system overload.†

and N_h is the number of non-friendly tracks scripted for the scenario. Tracks that become re-eligible for engagement due to an engagement failure or for other reasons are treated as completely new tracks, thus incrementing N_h.

DP will range from zero to 100. Timely engagement of all non-friendly tracks as they become eligible for engagement will result in a high score for DP. In these instances, the ratio TNE_i/TEE_i will be a small fraction (i.e., $TNE_i/TEE_i \rightarrow 0$), thus resulting in a small decrement to the operator penalty function. An operator will not, however, score 100 unless all non-friendly tracks are engaged instantaneously upon their declaration of eligibility, an unlikely outcome even under the automatic operating mode. A value of zero for DP will result from a situation in which no non-friendly tracks are engaged.

To illustrate the computation of DP, consider the situation illustrated in Figure 2-2. Ten tracks are scripted; all tracks are hostile. Tracks appear on the TBEQ at the times indicated and exit (i.e., are destroyed or become ineligible for engagement) at the times indicated; no engagement failures occur. The symbol "X" indicates the time when the operator engaged the track (i.e., Et).

†In PATRIOT, the operator does not actually fire a missile. Rather, by pressing "Engage" on the console assembly, he schedules a missile launch. The actual time of launch is determined by the system on the basis of available guidance resources and other considerations. Once the operator has scheduled a launch, his responsibility for a track has been discharged.

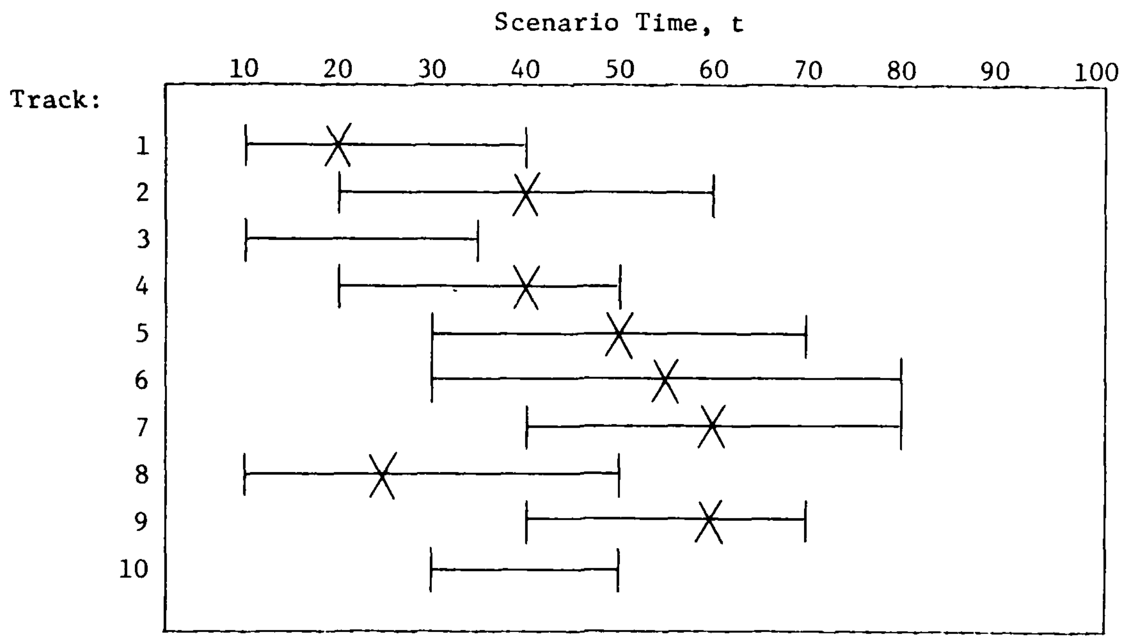

Figure 2-2. DP Track Times

Times from Figure 2-2 relevant to the computation of DP are summarized in Table 2-1.

Table 2-1. DP Summary Times

Track	EEt	Et	TNE	TEE	TNE/TEE
1	10	20	10	11.5	0.87
2	20	40	20	21.5	0.93
3	10	(35)†	25	25.0	1.00
4	20	40	20	21.5	0.93
5	30	50	20	21.5	0.93
6	30	55	25	26.5	0.94
7	40	60	20	21.5	0.93
8	10	25	15	16.5	0.91
9	40	60	20	21.5	0.93
10	30	(50)	20	20.0	1.00
					9.37

†The numbers in parentheses indicate that the track was not engaged.

Finally, DP is computed as

$$DP = \left[1 - \frac{\sum^{N_h} (TNE_i/TEE_i)}{N_h} \right] * 100 = \left[1 - \frac{9.37}{10} \right] * 100$$

$$= [1 - .937] \times 100$$

$$= (.063)(100) = 6.3$$

Defense of Assets. A primary objective of any air defense system is to prevent damage to friendly assets resulting from attack by hostile aircraft. Accordingly, an important index of operator performance is the extent to which protected assets are penetrated (i.e., exposed to risk).

DA is a measure of the number of physical asset penetrations by non-friendly aircraft weighted by the value of the penetrated assets. Quantitatively, DA is given as:

$$DA = \left[1 - \frac{\sum^{N_h} \sum AV_j \delta_{ij}}{\sum^{N_h} \sum AV_j \gamma_{ij}} \right] * 100. \qquad (2\text{-}2)$$

In expression (2-2), N_h is the number of non-friendly tracks scripted for asset penetrations; AV_j = (11.0-ATC_j), where ATC_j is the asset-threat category assigned to asset j;

$$\delta_{ij} = \begin{cases} 1 \text{ if asset } j \text{ is physically penetrated by track } i, \\ 0 \text{ if asset } j \text{ is not physically penetrated by track } i; \end{cases}$$

and

$$\gamma_{ij} = \begin{cases} 1 \text{ if asset } j \text{ is scripted for penetration by track } i, \\ 0 \text{ if asset } j \text{ is not scripted for penetration by track } i. \end{cases}$$

DA will occur in the interval zero to 100. An operator will receive a score of 100 if no assets are penetrated by non-friendly aircraft. If all scripted asset penetrations occur, then DA will equal zero.

To illustrate the computation of DA, consider the following three asset example:

Asset	ATC	AV
A	2	9
B	2	9
C	9	2

Scripted and actual penetrations are given as follows.

Non-friendly Track	Potential Asset Penetrations	Scripted Asset Penetrations	$AV_j\gamma_{ij}$	Actual Asset Penetrations	$AV_j\delta_{ij}$
1	A,B,C	A,B	18	A	9
2	A,B,C	C	2	C	2
3	A,B,C	A,B,C	20	A,B,C	30
4	A,B,C	A	9	None	0
5	A,B,C	A,B,C	20	B,C	11
			69		52

The resulting value for DA is computed as

$$DA = \left[1 - \frac{\sum^{N_h} \sum AV_j \delta_{ij}}{\sum^{N_h} \sum AV_j \gamma_{ij}} \right] * 100 = \left[1 - \frac{52}{69} \right] * 100$$

$$= [1 - .61] * 100 = .39 * 100 = 39.0.$$

Resource Utilization. Given that the PATRIOT system will likely be deployed in a theater where friendly forces are heavily outnumbered, it is essential that operators make efficient use of available defensive resources. RU is included in the performance evaluation scheme to assess the extent to which operators properly assign missiles to tracks as opposed to wasting them through actions such as requesting low probability engagements, assigning several missiles to the same track (i.e., engaging the same track more than once), and so forth.

Quantitatively, RU is defined as a function of the number of missiles defined as wasted versus the number of missiles properly assigned to threatening tracks:

$$RU = \left[1 - \frac{NMW}{(NML-NMF)} \right] * 100. \qquad (2-3)$$

In (2-3), NMW is the number of missiles wasted (improperly assigned, aborted, failed, and so forth) as a result of inappropriate operator actions;

NML is the total number of missiles launched;

and NMF is the total number of missile failures due to system fault (i.e., launch, guidance, or intercept failures) and not to operator actions.

RU will range over the interval

$$\frac{N_r}{(NML-NMF)} \leqq RU \leqq 100,$$

where N_r is the number of separate non-friendly tracks requiring engagement. The lower bound on RU recognizes that the system's internal logic will not permit all missiles to be wasted. If all missiles fail due to system fault, RU is arbitrarily set at 100. Also, if NML = 0, RU is arbitrarily defined to be 100.

As an example of the computation of RU, suppose that a total of 12 missiles are launched on 10 tracks. One of the missiles suffers an intercept failure, but the track is engaged again successfully. One track is inadvertently engaged twice. In this example, NML = 12, NMF = 1, and NMW = 1 resulting in an RU score of

$$RU = \left[1 - \frac{NMW}{(NML-NMF)}\right] * 100$$

$$= \left[1 - \frac{1}{(12-1)}\right] * 100$$

$$= [1 - \frac{1}{11}] * 100$$

$$= [1 - .09] * 100$$

$$= [.91] * 100 = 91.0.$$

Factor Combination. The three components characterizing operator performance at the mission-descriptive level are combined to form a single measure of effectiveness defining the system-descriptive level of the performance hierarchy. Functional forms judged appropriate for combining the three mission components into a single system-descriptive measure (the System Performance Measure, or SPM) fall into two categories representing opposite ends on a continuum of potential combination rules: an additive model and a multiplicative model. The additive combination rule takes the form illustrated below:

$$SP(A) = W_1 DP + W_2 DA + W_3 RU. \qquad (2\text{-}4)$$

In (2-4), the weights W_i represent the importance of each of the mission components to system performance (SP). If the W_i are selected such that they sum to one, then SP will occur in the interval 0-to-100.

A multiplicative combination rule takes the form illustrated in expression (2-5):

$$SP(M) = DP^{W_1} * DA^{W_2} * RU^{W_3}, \qquad (2\text{-}5)$$

where the W_i again reflect the importance of the individual mission-descriptive components to system performance. Again, if the W_i sum to unity, SP will occur in the interval 0-to-100.

In selecting an appropriate combination rule for SP, two characteristics of each of the aggregation models should be considered:

1. Model compensatoriness
2. Component independence

The additive combination rule is compensatory; that is, high scores on one or two of the mission-descriptive performance measures (denoted herein as Mission Performance Measures, or MPMs) will compensate for lower scores on one or more of the other MPMs. A multiplicative combination rule is less compensatory than an additive model; thus, a high score on one MPM does not compensate for lower scores on other components.

Considering the issue of component independence, the additive combination rule is appropriate when individual components are reasonably independent. That is, the additive model is appropriate in situations where the individual mission components address separate, independent aspects of system performance. Multiplicative aggregation models are suitable in situations involving correlated or "overlapping" components.

Both of these considerations suggest the use of the multiplicative combination rule in defining the SPM. First of all, poor operator performance

in any of the mission-descriptive areas can have disastrous consequences in the long run. Thus, since the real-world is not compensatory, so to speak, the system performance measure should not be either. This attitude is demonstrated in previous work of a similar nature done in the Air Defense community. In many of these studies, multiplicative combination rules have been employed, thus reflecting an opinion that low performance on one or more performance components should result in a lowered aggregate performance score (Howard, 1980).

Also to be considered in the selection of a combination rule is the fact that the MPMs comprising the SPM cannot be assumed to be independent. If an operator manages his airspace poorly (resulting in a lowered DP score), then there is a higher probability that assets under his protection will be physically penetrated. In such situations, the occurrence of launch, guidance, or intercept failures could result in asset penetrations that would not have occurred in a better managed situation in which the operator would have more time to recoup his defensive position.

To illustrate the computation of the SPM from the MPMs, the scores from the three examples cited previously are computed below. Although the multiplicative combination rule is judged more appropriate in the present situation, the additive rule is also illustrated. Values for W_1, W_2, and W_3 have arbitrarily been set at 0.4, 0.5, and 0.1, respectively. Recall that W_1 is the weight associated with DP, W_2 the weight associated with DA, and W_3 the weight associated with RU. Also recall that in the previous examples

$$\begin{aligned} DP &= 6.3, \\ DA &= 30.9, \\ \text{and } RU &= 91.0. \end{aligned}$$

Following these preliminaries, the multiplicative SPM is computed as:

$$\begin{aligned} SP(M) &= DP^{W_1} * DA^{W_2} * RU^{W_3} \\ &= 6.3^{.4} * 39.0^{.5} * 91.0^{.1} \\ &= 20.5 \end{aligned}$$

The score obtained using the additive combination rule is:

$$\begin{aligned} SP(A) &= W_1 DP + W_2 DA + W_3 RU \\ &= (.4)(6.3) + (.5)(39.0) + (.1)(91.0) \\ &= 31.1 \end{aligned}$$

Note that SP(A) is somewhat higher than SP(M), thus reflecting the compensatory nature of the additive combination rule.

Personnel Performance Measures

The lowest level in the performance hierarchy consists of the personnel performance criteria. These measures address performance on the individual tasks that comprise the PATRIOT console operator's job. As Meister (1976) notes, the personnel performance criteria address a very molecular level of operator performance such as reaction times, response accuracy, response variability, and the like.

Because of a number of practical limitations that will be discussed later in the report, it was necessary to restrict the consideration of personnel performance criteria to operator activities denoted as Air Defense Mission tasks. The individual tasks and associated task elements making up this set are listed as follows:

1. Prepare information displays for scenario.[†]

2. Observe displays and tracks prior to engagement.
   - Press Track Amplification Data
   - Press Clear Tab (optional)

3. Hook Tracks.
   a. Direct Cursor Hook:
      - Position Joystick
      - Press Hook
   b. Successive Proximity Hook:
      - Position Joystick
      - Press Hook (repeating this action constitutes a successive proximity hook)
   c. Numeric Hook:
      - Press Numeric Hook
      - Key digits (ID number for track being hooked)
      - Press Numeric Hook
   d. Sequential Hook:
      - Press Engagement Data
      - Press Sequential Hook
   e. Automatic Hook:
      - Press Alert Acknowledge [following Priority Engagement Alert (PEA)]

4. Engage Tracks.
   - Press Engage

5. Update To Be Engaged Queue.
   - Press Engagement Data
   - Press Clear Tab (optional)

6. Alert Responding.
   - Press Alert Acknowledge

[†]Task 1 is carried out through a special series of actions governed by Standard Operating Procedure (SOP) and is usually executed one time at the start of a scenario.

Task 3, Hook Tracks, presents the operator with a decision-making situation. Given that a track is to be hooked, one of the five hook modes is employed to that end. Under various conditions, however, certain hooking modes may be more appropriate than others. In order to clarify this issue, Table 2-2 displays the conditions under which a hook can be made and the appropriate hooking mode (or modes) for each condition.

Table 2-2. Hook Mode Contingencies

	Hooking Mode				
Condition	Direct Cursor	Successive Proximity	Sequential	Numeric	Automatic
Search/Observation Engagement	X(2)	X(3)	N	X(1)	N
TBEQ Engagement	I	I	X	I	N
Priority Engagement Alert (PEA) in effect	I	I	I	I	X
Track Re-engagement: (a) PEA	I	I	I	I	X
(b) No PEA	I	I	X(1)	X(2)	N

X - Appropriate Mode. Number in parentheses indicates preferential order.

N - Not possible. Attempt constitutes an operator error.

I - Inappropriate mode (but possible).

The tasks and task elements listed above constitute the set of operator actions defining the personnel performance level of the performance hierarchy. Prior to moving from the "what is done" stage to the specification of performance criteria, it is necessary to determine what individual operator actions are appropriate (or admissible) under various conditions. A portion of the resolution of the issue of what is appropriate is outlined in Table 2-2, Hook Mode Contingencies. Table 2-2 does not, however, provide a complete resolution of the issue.

Operator actions in the PATRIOT system are precipitated by two classes of stimuli: system cues and previous operator actions. Table 2-3 provides a list of system cues and admissible following responses. System cues 5 through 8 in Table 2-3 are not explicitly treated in the current effort, thus admissible operator responses to these cues are not listed. To complete the specification of admissible following responses, Table 2-4 presents the response contingency matrix for the second class of precipitating stimuli, previous operator response.

Table 2-3. System Cues and Admissible Operator Responses

System Cue	Admissible Operator Responses
1. Track on display on scope	•Position Joystick •Press Numeric Hook •Press Engagement Data
2. Alert Message Line	•Press Alert Acknowledge
3. Blinking Track Number (in TBEQ)	•Press Engagement Data
4. Targets in TBEQ	•Position Joystick •Press Numeric Hook •Press Engagement Data •Press Sequential Hook
5. Fire Unit Commander Audio Command	•To Be Determined (TBD)
6. Battalion Commander Audio Command	•TBD
7. Adjacent Fire Unit Communication	•TBD
8. Hardware Fault	•TBD

Table 2-4

Admissible Following Responses
for PATRIOT Console Operators

(Read Row by Column)

Action \ Following Action	Position Joystick	Press Hook	Press Engage	Press Numeric Hook	Key Digit	Press Alert Ack.	Press Eng. Data	Press Seq. Hook	Press Cancel Hook	Press Clear Tab	Press Trk. Amp. Data	Prepare
Position Joystick	A	A	E	C_1	E	C_2	C_3	C_4	X	X	E	E_1
Press Hook	A	A	A	C_1	E	C_2	A	C_4	X	X	A	E
Press Engage	A	C_5	E	A	E	A	A	A	X	X	A	C_8
Press Numeric Hook	C_6	A	E/A	E	A/E	C_2	X	C_4	X	X	X	C_8
Key Digit	C_6	A	E	E/A	A/E	C_2	X	C_4	X	X	X	C_8
Press Alert Ack.	A	X/E	A	A	E	A	A	A	X	X	A	C_8
Press Eng. Data	A	E	E_1	A	E	A	A	A	X	X	X	C_8
Press Seq. Hook	A	E	A	C_1	E	A	A	C_7	X	X	A	C_8
Press Cancel Hook	X	X	X	X	X	X	X	X	X	X	X	C_8
Press Clear Tab	X	X	X	X	X	X	X	X	X	X	X	C_8
Press Trk. Amp. Data	A	A	A	E/A	E	A	A	E	X	X	A	C_8
Prepare	A	E	E	A	E	A	A	A	X	X	C_9	X

Key:

A - Allowable transition
C - Conditionally allowable transition
E - Error
X - Not meaningful, but not an error

Note: Subscripts on C or E entries are explained in Appendix A.

The information contained in Tables 2-2, 2-3, and 2-4 provides the basis for the development of personnel performance criteria. From these tables, a scoring system for individual operator response protocols was developed. The scoring system provides: (1) A characterization of responses as admissible or inadmissible and, in selected cases, the identification of admissible, but less-than-optimal responses; (2) reaction times for system cues; (3) lag times in response-response sequences; and (4) response-response and system cue-operator response conditional probabilities (these data address Meister's issue of response variability).

The Task Performance Measures (TPMs) obtained as described above serve two objectives. Objective one concerns the development of rational, empirically-based performance standards for the PATRIOT console operator Military Occupational Specialty (MOS) (i.e., 24T). Using the TPM data in combination with the MPMs and the SPM, it will be possible to provide valid SOPs for the operator's job. Minimum time requirements for task completion based on actual, as opposed to assumed, human operator capabilities will also be available (Hoffer & Howard, 1979).

The second objective served by the collection of task performance (TP) data involves parameterizing the PATRIOT operator optimization model. In order to develop the optimization model, it is necessary to characterize operator performance in terms of response latencies and task completion times. The application of TP data in addressing this objective is described in additional detail in section 3 (Model Parameterization).

Situational Difficulty: A Critical Moderator of Operator Performance

The Problem

An issue associated with using the SPM and MPMs to evaluate individual operators or crews concerns the equivalence of scores across evaluation scenarios having varying levels of difficulty. For example, if two operators both achieve an SP score of 80, but the individual scores are obtained under different evaluation scenarios, are the two operators to be considered comparable in their performance? The obvious answer to this query is, "not necessarily."

When comparing the performance of console operators acting against different threats, allowance must be made for characteristics of the engagement environment and the threat situation that moderate performance. Observation of only raw operator performance indices can be misleading since doing so fails to consider the differential nature of the task demands (i.e., operational environment) placed upon the operators. In accord with this

view, a corollary to the task of developing a series of operator performance measures is the development of an index that can be used to adjust raw SP/MP scores to reflect the difficulty of the operational environment in which the data were obtained.

As part of a related effort (preceding the current project), a preliminary PATRIOT scenario difficulty index was developed and evaluated (Hosch, Starner, & Howard, 1980). This measure of scenario difficulty (denoted herein as the UTEP metric), is based upon the performance of the PATRIOT system in automatic mode. The UTEP metric is, essentially, the reciprocal of automatic system performance. If the system in automatic mode has a difficult time coping with a scenario, as indexed by a lowered system performance score, then the scenario is difficult, and vice versa.

In application, the UTEP difficulty metric provides results that are apparently reasonable. For example, Spearman's rank correlation coefficient between UTEP scenario difficulty results on a series of test scenarios and difficulty rankings provided by subject matter experts (SMEs) from the Air Defense School at Ft. Bliss was r = 0.96.

Development of a Situational Difficulty Index

In light of the criticisms of the UTEP metric as an index of situational difficulty (SD), a decision was made to develop a situational difficulty index (SDI) that is not subject to the same limitations. At the start, the position was taken that an ideal SDI should be: (1) a priori, (2) not confounded with operator and/or system performance, and (3) developed from a solid human factors perspective. The term a priori, in this context, denotes a measure that can be computed in advance of, and in isolation from, operator or system actions.

After a review of previous work concerned with the quantification of situational difficulty in a human-machine environment (e.g., Conrad, 1956; Siegel & Wolf, 1969; McCormick, 1976; Hosch, Starner, & Howard, 1980; Siegel & Federman, 1980; Swain & Guttman, 1980), it was judged reasonable to base the SDI on the concept of operator load stress. Following Conrad (1956), load stress is characterized as a function of the product of load and speed. Operator load is defined as the variety (i.e., type and number) of stimuli to which an operator must attend; speed is alternately defined as the number of stimuli occurring per unit of time (i.e., rate of change), or the time available to process each stimulus (McCormick, 1976).

Given the conceptual orientation expressed above, the next step in the development of the SDI involved defining load stress in terms of observables within the PATRIOT human-machine environment. After an analysis of the operational environment, operator load stress at time $\underline{t}$ [denoted as $\Phi(t)$] was characterized as a function of the following aspects of the engagement environment:

1. The total number of tracks on the situation display, n;
2. The number of hostile tracks, h ($h \leqq n$);
3. The number of unknown tracks, u ($u \leqq n$);
4. The mean velocity of hostile and unknown tracks, $\overline{V}$. Track velocity is a correlate of speed (as defined by Conrad).

Specifically, $\Phi(t)$ is defined as

$$\Phi(t) = n(t)*h(t)*\overline{V}[h(t)]*u(t)*\overline{V}[u(t)], \qquad (2\text{-}6)$$

where $n(t)$ is the total number of tracks on the situation display at time $\underline{t}$;

$h(t)$ is the total number of hostile tracks at time $\underline{t}$, $h(t) \leqq n(t)$;

$\overline{V}[h(t)]$ is the mean velocity of the $h(t)$ hostile tracks at time $\underline{t}$;

$u(t)$ is the number of unknown tracks at time $\underline{t}$, $u(t) \leqq n(t)$;

and $\overline{V}[u(t)]$ is the mean velocity of the $u(t)$ unknown tracks at time $\underline{t}$.

Total situational difficulty for a given scenario is obtained by integrating $\Phi(t)$ over scenario time. That is,

$$SD(R) = \int_{t_s}^{t_e} \Phi(t)dt, \qquad (2\text{-}7)$$

where SD(R) denotes raw situational difficulty and t_s and t_e are the scenario starting and ending times, respectively.

Since it was judged desirable that the SDI be dimensionless and in the interval [0,1], (2-7) is normalized to reflect a worst case situation:

$$SD(N) = \frac{n_t}{(N_{max})^2 (h_{max}) (u_{max}) (\overline{V}_{max})^2 (t_{max})} \int_{t_x}^{t_e} \Phi(t)dt. \qquad (2\text{-}8)$$

In (2-8), SD(N) denotes normalized situational difficulty;

n_t is the total number of tracks scripted for the scenario;

N_{max} is MAX(n_t) across evaluation scenarios;

h_{nax} is MAX[h(t)] across scenarios;

u_{max} is MAX[u(t)] across scenarios;

$\overline{V}_{max}$ is MAX[MAX[$\overline{V}$(h(t))], MAX[$\overline{V}$ u(t))]]

and t_{max} is the maximum engagement length across scenarios; i.e., t_{max} = MAX(t_e-t_s).

The application of the normalization constant scales the SDI to fall between zero and one. An SD value of "0" indicates no activity (i.e., load stress) and a value of "1" indicates the most difficult scenario constructed to date.

To illustrate the computation of the SDI, consider the scenario illustrated in Figure 2-3. In this scenario, ten tracks are scripted, all are hostile; $\overline{V}$[h(t)] = 300 meters per second (m/sec.) (constant); $\overline{V}_{max}$ = 750 m/sec. The tracks appear on the *situation* display (and on the TBEQ) at the times indicated; the tracks exit (i.e., become ineligible for engagement) at the times indicated. For purposes of illustration, it is assumed that N_{max} = 15 tracks, h_{max} = 15 tracks, and t_{max} = 100 seconds.

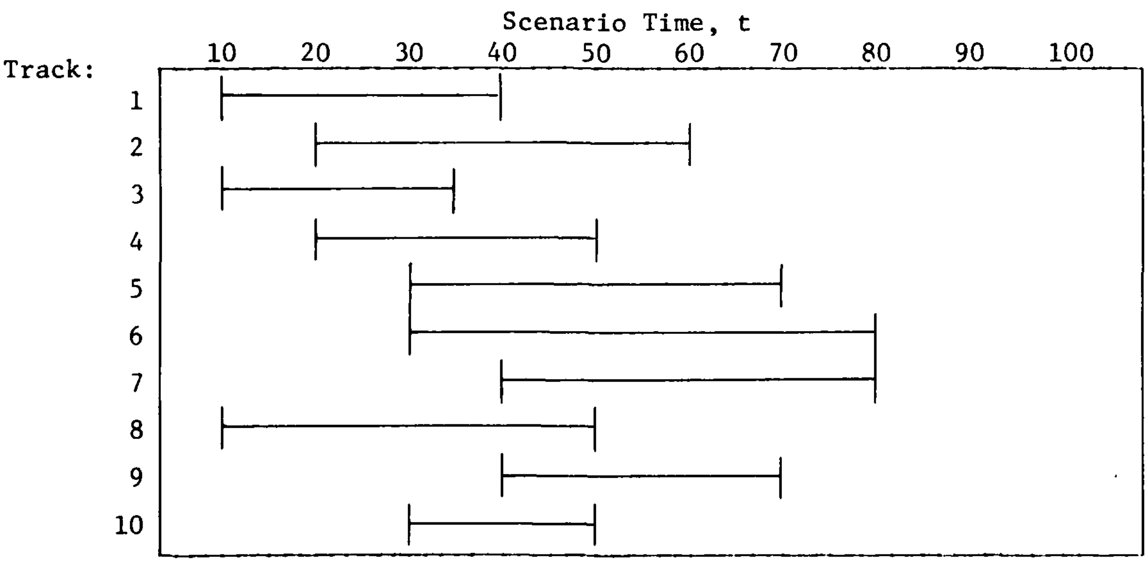

Figure 2-3. SDI Track Times

A plot of $n(t)$ [(which is also equal to $h(t)$] as a function of scenario time is presented in Figure 2-4. The value of SD(R) [expression (2-7)] is obtained by integrating $\Phi(t)$ (i.e., 750m/sec. times the squared values of the function plotted in Figure 2-4) from $t_s = 10$ to $t_e = 80$. Performing this integration, the following value is obtained:

$$SD(R) = \int_{10}^{80} n(t)h(t)\bar{V}[h(t)]dt = 5{,}985E3^{\dagger}.$$

The normalization constant is computed as follows:

$$\frac{n_t}{(N_{max})^2(h_{max})(\bar{V}_{max})(t_{max})} = \frac{10}{(225)(15)(750)(100)} = 3.9506E\text{-}8.$$

Finally, SD(N) is computed as:

$$SD(N) = 5985E3 * 3.9506E\text{-}8 = .2364.$$

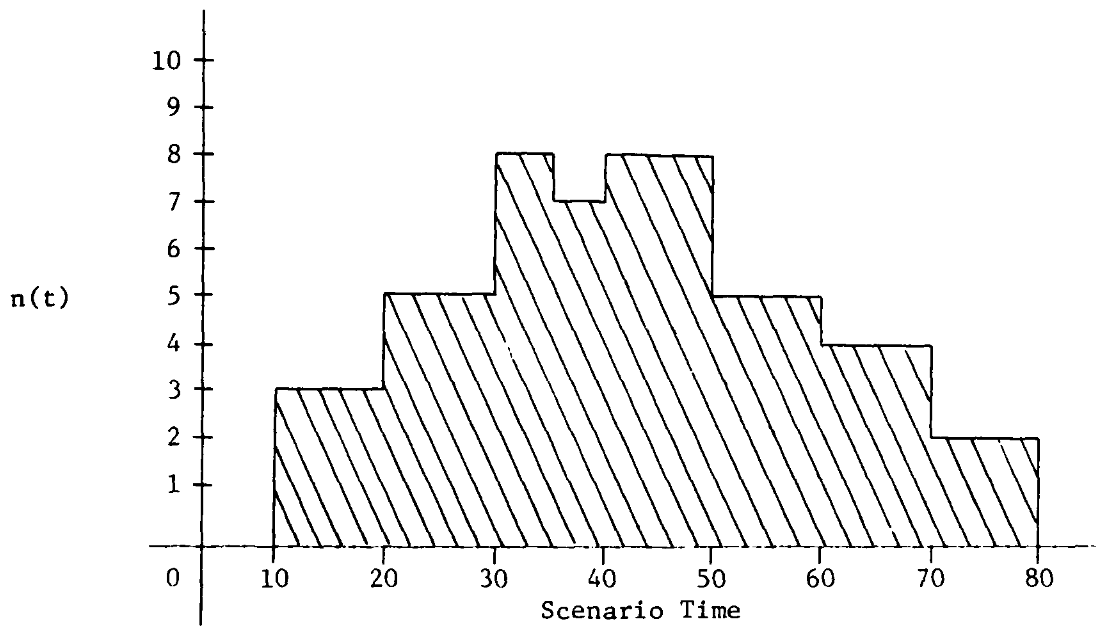

Figure 2-4. Tracks as a Function of Scenario Time

† In an application where any of the members of $\Phi(t)$ is zero [e.g., $u(t)$], the zero member is arbitrarily assigned a value of unity.

As indicated previously, SD(N) is to be used to adjust raw SP/MP scores to reflect the differential nature of the conditions presented to an operator during an evaluation scenario. The measure is an index of the load stress by time placed on an operator during an engagement. Implied in its definition is the view that difficult scenarios are characterized by high load stress over a protracted time period. Easier scenarios have either lower load stress, in general, or periods of high load stress separated by periods with little or no activity (i.e., engagement "bursts"). The normalized, time-average value of $\Phi(t)$ defines total situational difficulty.

Application of the SDI

The final theoretical issue under the topic of performance assessment concerns the application of the SDI in adjusting SP and MP scores. In order to apply the SDI, however, it is first necessary to compute it. Operators interact with scenarios either to increase or to decrease de facto difficulty levels. Hence, each operator, in essence, faces a unique evaluation situation. As noted previously, what is desired on the part of the SDI is an a priori, standard index of situational difficulty that can be obtained independent of operator or system performance. In practice, such an index can be obtained in one of two ways: (1) by running scenarios in automatic mode and recording SD, or (2) by running scenarios in semi-automatic mode with no operator intervention and noting SD. The former method will result in a theoretical and practical lower bound for the SDI; the latter will provide a theoretical and practical upper bound for the SDI. A problem, however, is determining which value is appropriate for use in adjusting operator scores. Since a satisfactory resolution to this issue was not apparent going into the project, a decision was made to obtain both values and then to determine empirically which index apparently works best in adjusting SP/MP scores.

Following the computation of the SDI, a second issue concerns a method of applying it to obtain adjusted SP/MP scores. After debating the pros and cons of several usage procedures (e.g., direct multiplication of scores by the SDI), a decision was made to approach the adjustment problem following an analysis of covariance (ANCOVA) framework [see Bock (1975) for a discussion of the analysis of covariance as applied to behavioral research]. The ANCOVA approach to the adjustment problem proceeds as follows. First, the form of the relationship between the SPM/MPMs and the SDI is established. This is done through a regression analysis of the relationship between scenario SDI values and SP/MP results for a range of representative PATRIOT console operators (i.e., a standardization sample). The results of a hypothetical regression analysis are as depicted in Figure 2-5. The form of the regression line can be any order polynomial. For exemplary purposes, however, the form of the regression shown in Figure 2-5 is linear.

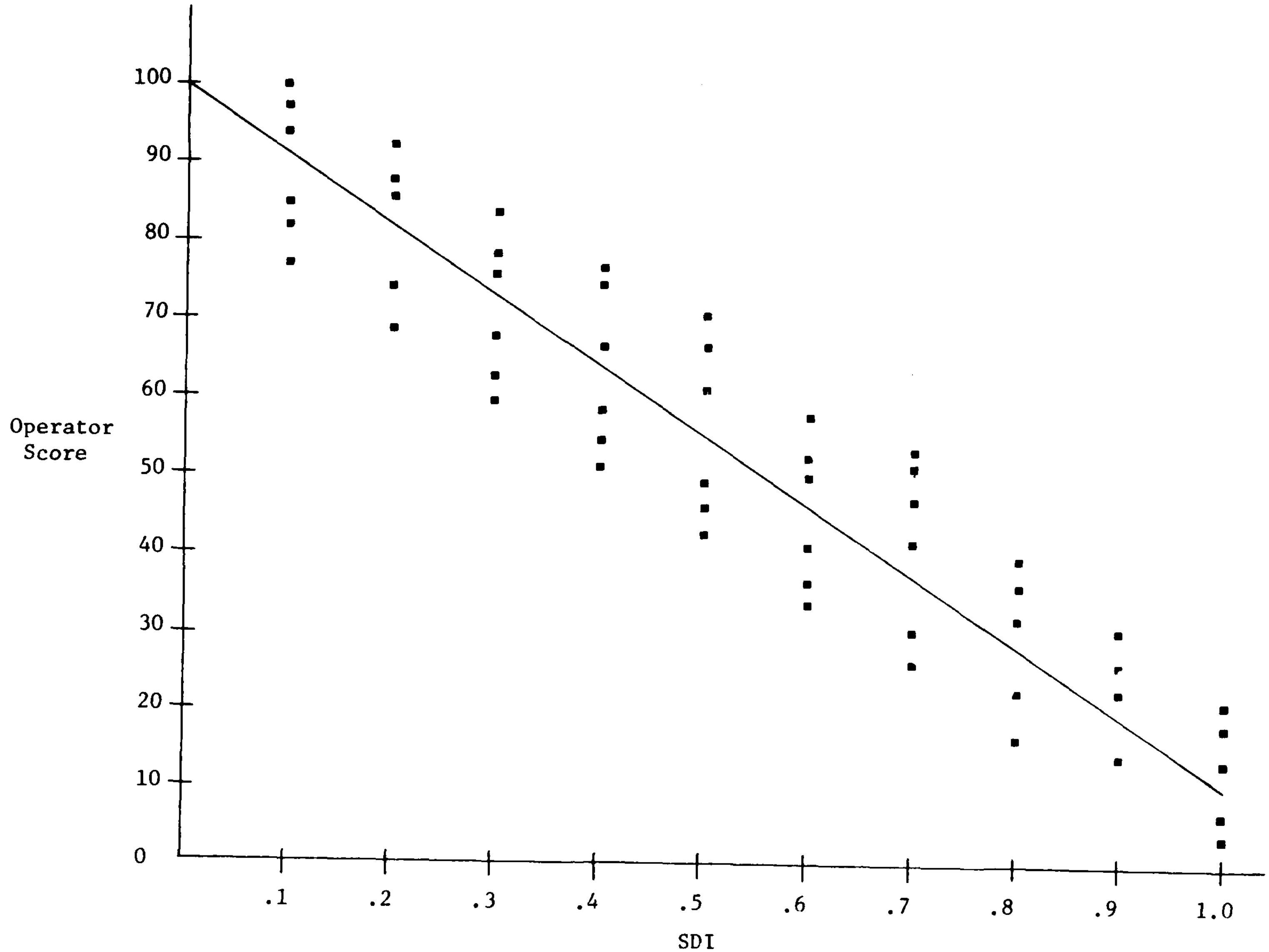

Figure 2-5. Hypothetical SDI-SPM/MPM Regression Results

In most applications, there will be considerable dispersion of operator scores about the regression line. This situation is to be expected and actually constitutes the key to the remainder of the adjustment procedure. The predicted scores, $\hat{Y}$, are estimates of the mean operator score given a particular level of difficulty; that is, $E(\hat{Y}) = \mu_{y \cdot SD}$. For example, consider a situation in which the regression equation is established as

$$\hat{Y} = 100 - 90(SD).$$

Using this regression equation, the expected operator score for a scenario having an SD value of 0.5 is $\hat{Y} = 55$.

After obtaining expected performance scores, the second step in the application of the SDI in obtaining difficulty-adjusted SP/MP scores is standardization. Following the ANCOVA framework, standardization is not carried out using raw SP/MP scores; rather, it is performed using the residuals obtained by subtracting expected operator scores from observed operator scores:

$$R = Y - \hat{Y}.$$

The expected value of R is zero. A positive residual indicates that an operator performed better than expected; a negative residual indicates that an operator performed poorer than average. How much better or worse is established through a reference to established operator norms (see Angoff, 1971).

To illustrate how such norms might be established and used, again consider the example situation. Assume that a hypothetical operator achieves an SP score of 72 on a scenario for which the expected score is 55. Now, also assume that the distribution of residuals about the regression line is normal with a standard error of estimate $\sigma_{y \cdot x} = 15$. An operator score of 72 thus represents a Z-score of

$$Z = \frac{X - \mu_{y \cdot x}}{\sigma_{y \cdot x}} = \frac{75 - 55}{15} = \frac{17}{15} = 1.13.$$

A Z-score of 1.13 indicates that the hypothetical operator scored at the 87th percentile relative to the standardization sample.

To aid in the derivation of normative results, field evaluation personnel can be provided with a nomograph for obtaining standard scores from difficulty-adjusted SP/MP scores. As an example, the nomograph that would apply in the previous example is provided as Figure 2-6. To use the nomograph, field personnel would merely enter the chart at the appropriate point on the ordinate (e.g., 1.13), interpolate across to the curve, and then down to the abscissa. The point at which the vertical line intersects the abscissa is the operator's standard score, in this case a percentile score of 87. For operator certification purposes, minimum acceptable performance levels can be indicated on the nomograph, as shown in Figure 2-6.

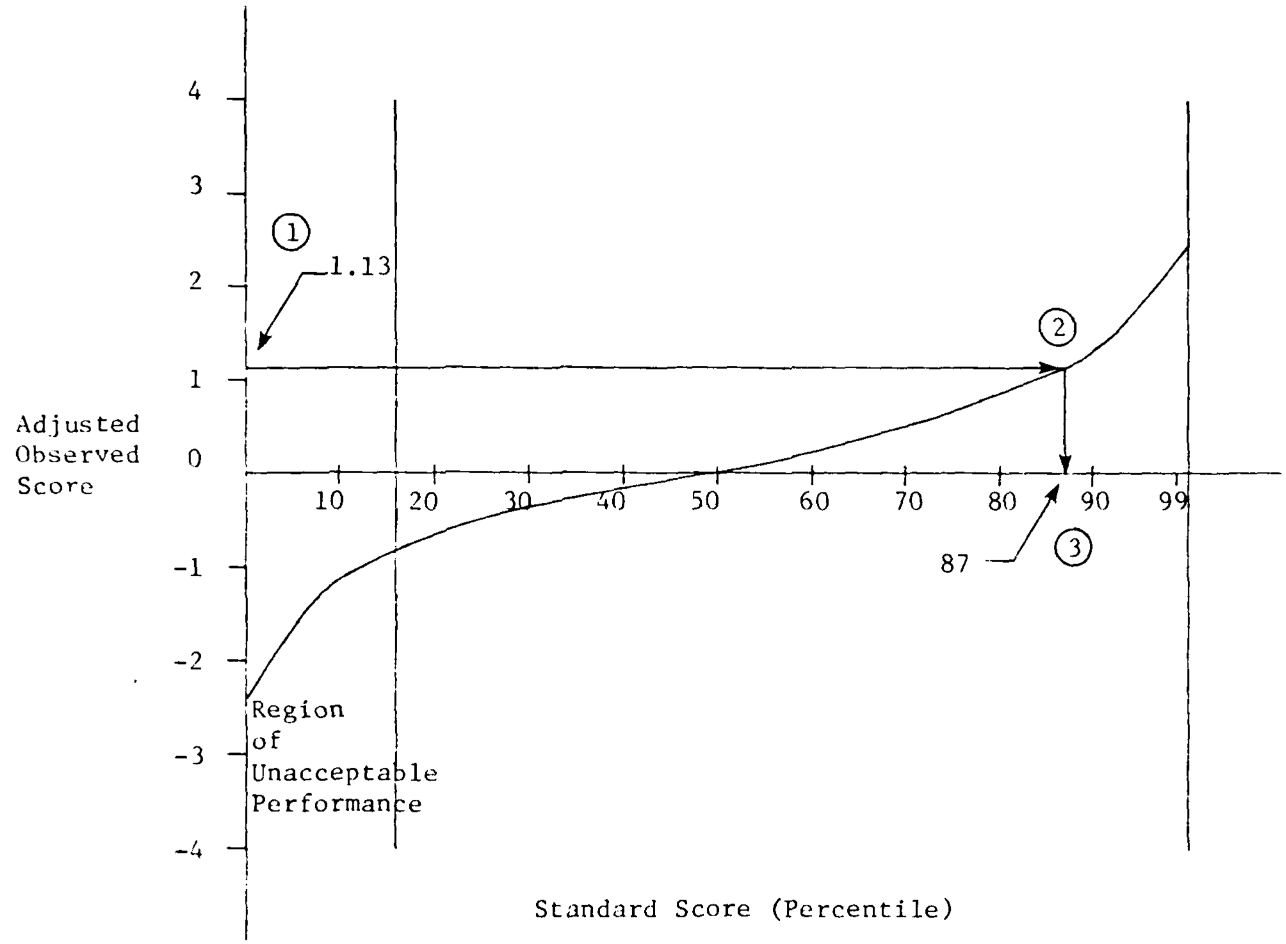

Figure 2-6. Exemplary Nomograph for Obtaining SP/MP Standard Scores

Implementation of the Performance Assessment Capability

The intention going into the project was to employ the operator performance assessment capability in support of the development of the operator optimization model. Specifically, SP/MP scores were to define PATRIOT operator performance (i.e., the criterion for optimization) and selected TP data were to be used to parameterize the optimization mode. Required SP, MP, and TP data were to be derived from test operator performance protocols obtained as part of a previous ARI in-house project, the PATRIOT Console Operator Performance Analysis (PCOPA) (see Howard, 1978, 1979a, 1979b). In this earlier project keypress/switch action responses and response time data were obtained from 67 test operators on 48 scenarios. A variety of supporting psychological, psychomotor, and biographical information on the test operator subjects was also collected.

Since the real system was not available for experimentation during Howard's study, test operator performance protocols were obtained using the PATRIOT Tactical Operations Simulator/Trainer (TOS/T) located at the U.S. Army Air Defense School Directorate of Combat Developments (USAADS DCD). The TOS/T is an environmental, full-task simulator for the PATRIOT system. It is designed to simulate the two-person ECS environment. However, at the time the test operator data were collected, the TOS/T was configured to handle a single operator only. As an additional constraint, only the Air Defense Mission tasks (listed previously) were enabled on the simulator. Thus, the PCOPA study addressed only these tasks.

After the operator performance assessment scheme was defined, it was discovered that the test operator protocols obtained in the PCOPA project did not contain all of the information necessary to compute SP/MP/TP scores. The test operator protocols contain switch actions/keypresses and their times of occurrence, but system cues were not recorded. This problem necessitated a different approach to scoring the test operator data than had been originally intended. Initially, the research plan called for scoring the protocols without additional access to the TOS/T. However, the lack of system cue data necessitated having to rerun each test operator protocol using the TOS/T and overlay system cue times on the original performance record. It was assumed that this procedure would present no problems; the replay, system cue overlay, and scoring process were to be done in a single pass through the PCOPA test operator data.

In accord with the original research plan, the project staff set about the task of modifying the TOS/T simulation software to: (1) replay test operator protocols, (2) overlay system cues, and (3) obtain SP/MP/TP scores. At first, the software modifications progressed smoothly and the required playback capability was developed ("Optimizing Operator Performance," 1982). The project staff then set about developing the scoring code. This, too, was partially developed and debugged without undue difficulty.

Upon the initial tryout of the SP/MP portion of the scoring program, an additional problem became apparent: The time "hacks" recorded on the PCOPA test operator performance records are absolute clock times, as opposed to relative TOS/T simulation battle times. Clock time, in this context, refers to absolute time external to the TOS/T simulation. Battle time, on the other hand, is relative and internal to the TOS/T simulation. Adding the SPM/MPM code to the TOS/T simulation software results in a noticeable slowing of simulation battle time relative to clock time. Hence, the clock time increments on the test operator protocols are no longer synchronized with battle time increments in the modified TOS/T simulations (i.e., simulations run with the scoring code included). As a result, the simple replay, overlay, and scoring of the test operator protocols was no longer possible.

Two solutions to the time synchronization problem were apparent. The first solution involved rerunning each of the 67 * 48 = 3216 test operator protocols without the scoring code in place and recording battle times instead of clock times. System cue data would also be included this second time around. The new performance protocols obtained in this fashion would then be scored directly, as originally planned. This first solution was rejected out of hand as being too costly. Assuming an average of 15 minutes to rerun each protocol, an estimated 3216 * 15 minutes = 803.75 hours = 33.49 days of continuous running on the TOS/T would have been necessary just to create the modified performance protocols. Given USAADS DCD's other committments, the required TOS/T time was not available.

The second potential solution to the time synchronization problem involved employing a software routine to "freeze" simulation clock time whenever the scoring code was activated. It was thought that such an approach would very nearly synchronize the recorded clock times with new simulation battle times. Having decided that this solution was feasible, development of the scoring capability again proceeded.

After adding the TPM code to the SPM/MPM code, an additional problem arose. The complete performance assessment package was too large to integrate readily into the existing TOS/T simulation software. It thus became necessary to separate the TP code from the SP/MP code and to develop a separate TP evaluation capability. This task was accomplished and the results are also documented in Optimizing Operator Performance on Advanced Training Simulations: Program Documentation.

Having solved the problems noted above, it was thought that the scoring of test operator protocols could finally commence. However, about this time the TOS/T began experiencing severe system hardware problems. The difficulties were of such a nature that it was not possible to determine whether the scoring software add-ons or the TOS/T itself was at fault. After a review of the situation by the project staff, USAADS DCD personnel, and the Contract Monitor, it was decided that the only reasonable course

of action under the conditions encountered was to bring all required software to a status from which implementation could easily take place once the TOS/T problems were resolved. As a result of this decision, planned test operator scoring did not occur. The consequences of this outcome also impacted significantly upon the development of the operator optimization model, as discussed in greater detail in section 3.

This section of the report has presented material relevant to the development of a performance assessment capability for PATRIOT ECS console operators. The capability is useful in itself, but, within the scope of the current effort, the objective of developing the scoring capability is to provide the criterion context within which to examine the relationship of factors such as selection, training, human-system integration, and doctrine with PATRIOT human-machine performance.

3. THE PATRIOT OPERATOR SIMULATOR AND SYSTEM UTILIZATION MODEL

The third objective of the current project concerned developing "models for predicting operator effectiveness relative to psychological data and psychomotor responses." As noted in section 1, the rationale behind the model development effort is to provide a means for optimizing PATRIOT man-machine system performance through:

1. The development of appropriate operator selection strategies;

2. The construction of a rational, progressive operator training program;

3. The study of human-machine integration issues such as operator-operator or operator(s)-computer function allocation;

4. The development of doctrine for the optimal employment of the system (e.g., deployment, operation, etc.).

Note that these four areas of concern address, first, the human component of the human-machine system, (i.e., selection and training); next, the interface of the human component with the machine component; and, finally, the total system's application.

Upon first review, the objective of developing a model, or series of models, for predicting operator performance from operator psychological and psychomotor characteristics would seem to be a relatively straightforward undertaking. Having available suitable measures of operator performance, it is necessary to: (1) obtain representative test operator performance protocols, (2) score the performance protocols, and (3) use least squares procedures to obtain the best regression equation for relating performance to operator characteristics. Consider, however, the true extent of the performance prediction problem. The objective is to derive a prediction model relating at least four major classes (corresponding to the four areas of concern noted above) of operator/situational variables to PATRIOT human-machine performance. In regression notation, the general form of this model is expressed as follows:

$$\underline{P} = f(\underline{\Psi},\underline{T},\underline{S},\underline{D}). \tag{3-1}$$

In (3-1), $\underline{P}$ represents a vector of dependent, performance-related variables;

$\underline{\Psi}$ represents a vector of psychological/psychomotor variables relevant to operator selection;

$\underline{T}$ represents a vector of independent variables relevant to training-related issues:

$\underline{S}$ represents a set of variables characterizing system conditions of interest (e.g., a particular function allocation scene);

$\underline{D}$ represents a vector of doctrinally-important independent variables;

and $f(\cdot)$ represents a polynomial combination function.

When all of the main effects and possible interactions in the full model (3-1) are considered, a standard experimental design approach to meeting the objectives of the SOW becomes questionable. Currently, it is practically impossible to obtain the numbers of trained test operator subjects necessary to collect sufficient data to provide reliable estimates of the regression parameters involved.

An alternative approach to studying all of the independent variables noted above as a set is to study the variable sub-sets individually (possibly over a longer period of time) and then to combine the results across sub-sets in order to make inferences about relationships in the system as a whole. Such an approach assumes, however, that the nature of the interactions among the variable sub-sets is known and can be quantified (see Meister, 1971). In the present situation, this assumption is probably untenable.

A second potential approach to the problem of developing a performance prediction model suitable for meeting the objectives of the project involves the development of a structural analog of the PATRIOT console operator. This model would be used to simulate operator behavior in the human-machine system. Such a model could serve as a partial surrogate for experimentation with real operators. It is not intended that such a model completely replace the study of actual operators. Rather, the simulation model would be used as a preliminary evaluation, or "screening", tool to provide the insight required for the design of efficient, more definitive studies involving real-world console operators. Given the problems involved in obtaining, training, and evaluating actual test operators, a decision was made to pursue this second approach to the PATRIOT human-machine performance enhancement problem.

General Modeling Approach

Formulating simulation models is typically more difficult than developing other kinds of quantitative models. For example, if a system can be described in a way that meets the assumptions of linear programming, there is an accompanying theory of model building that provides a guide to subsequent activities (e.g., Gillett, 1976). The theory describes the treatment

of data and outlines procedures for analyzing data to solve the problem. However, when a decision is made to employ a simulation model, few guidelines are available. This is because no well developed situation-independent theory of simulation, analogous to that of linear programming, currently exists (Emshoff & Sisson, 1970). Most treatments of the topic area "system simulation" present a series of guidelines illustrated with case studies (e.g., Shannon, 1975; Bobillier, Kahan, & Probst, 1976).

As simulation techniques have become more refined, a methodology of simulation has begun to appear (Fishman, 1978; Law & Kelton, 1982). Systems that consist of discrete units (e.g., job lots, customer arrivals, job tasks, decision processes, etc.) flowing in a sequence (e.g., a machine tool center, a bank, a PATRIOT ECS, etc.) are generally representable in a common form. The key to the simulation of such discrete-entity systems is a means of representing processes. In simulation terminology, a process is an activity that proceeds over time. The initiation, modification, or termination of the process is referred to as an event. When an event occurs, the overall state of the system changes. In a discrete event simulation, processes are not modeled explicitly. Rather, processes are simulated by modeling the events that affect their status.

Following this general methodology, the first activity in developing a simulation model for a discrete event system is to construct an event list. The event list contains a description of all events that will occur at some time during the simulation. During the course of the simulation, the event list also specifies the times at which various events are scheduled to occur. The event list is augmented as additional events are scheduled and the system state changes. Event and time prediction routines thus become central elements of a discrete event simulation model. These routines specify how the system and its environment determine events and event time durations.

Once an event list and an event prediction and timing mechanism have been specified, the basic flow of a discrete event simulation is as portrayed in Figure 3-1. The simulation sequence begins when the event list is queried to find the first event. The simulation clock is then advanced to that time. The next step in the process is to determine the class of event that has occurred. Is it the completion of a job, the arrival of an order, or the appearance of a track in the TBEQ? Next, primary event subroutines are used to change the system's status. If required, conditional event routines are used to indicate events resulting from the new system state. When all conditional events have been added to the event list, the entire process is repeated. The repetition of the process "moves" the system simulation through time, so to speak. As the simulation proceeds, system performance is recorded by tabulating appropriate indices.

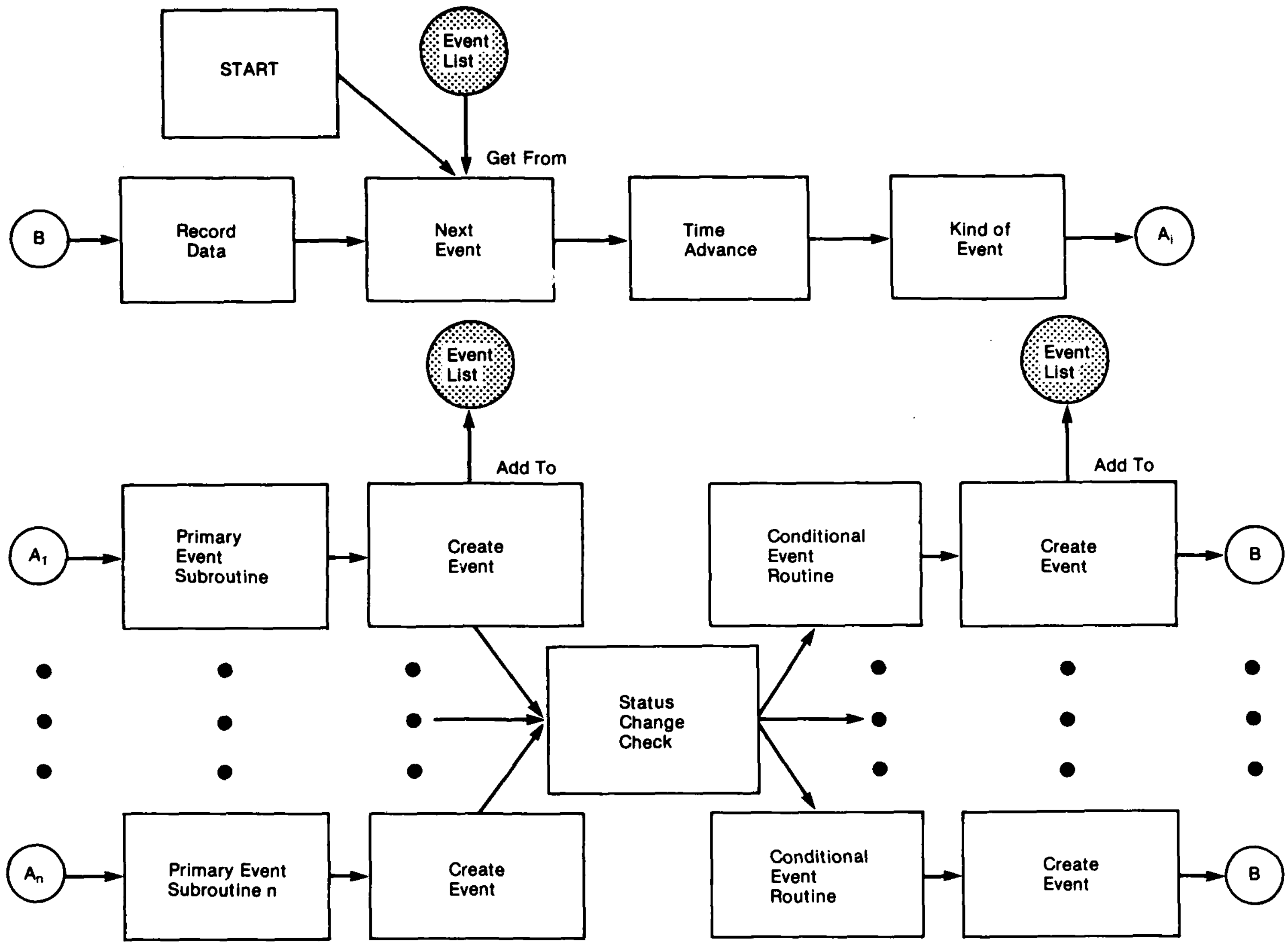

Figure 3-1. A Basic Structure for Discrete Event Simulation (Adapted from Emshoff and Sisson, 1970)

Model Structure and Development

Following the general description of a discrete event simulation presented in the previous paragraphs, the first activity in developing a simulation model of a PATRIOT ECS console operator is to assemble an event list. After listing the events of interest, the second activity is to devise an event generation and timing capability (i.e., to develop a procedure for creating events and specifying event times). The event generation capability is based upon a logical model characterizing the behavior of the real system (i.e., the behavior of an actual console operator).

As noted in section 2, the current project is restricted to the study of the six Air Defense Mission tasks listed below:

1. Prepare information displays for scenario.
2. Observe displays and tracks prior to engagement.
3. Hook tracks.
4. Engage tracks.
5. Update TBEQ.
6. Alert responding.

With the exception of Task 1 (which represents a special case), operator tasks of interest are accomplished through the execution of a finite set of switch actions and key presses, listed as follows:

1. Position Joystick
2. Press Hook
3. Press Engage
4. Press Numeric Hook
5. Key Digit
6. Press Alert Acknowledge
7. Press Engagement Data
8. Press Sequential Hook
9. Press Cancel Hook
10. Press Clear Tab
11. Press Track Amplification Data

For example, the task "Hook Tracks" using the Numeric Hook mode is carried out through the following series of discrete operator actions:

1. Press Numeric Hook
2. Key digit
3. Key digit

   Track number for track being hooked

4. Press Numeric Hook

A hooked track is then engaged by pressing the "Engage" key on the console assembly. Since they constitute the universe of interest, the switch actions/key presses listed above comprise the event list for a PATRIOT ECS console operator simulation model (denoted herein as the PATRIOT Operator Simulator and System Utilization Model, or POSSUM).

Having specified the elements of the event list, the second step in the development of an operator simulation model is to devise an event scheduling and timing procedure. From section 2, recall that operator actions are prompted from one of two sources: system cues or the last action taken by the operator. System cues appear on the display console and, for the present, are represented by the following stimuli:

1. Track on display on scope
2. Alert message line
3. Blinking track number
4. Target(s) in TBEQ

Lawful operator responses to system cues are outlined in Table 2-3. Operator response-response contingencies are provided in Table 2-4. Taken together, these two sources define the logical basis for the operator model.

A review of Tables 2-3 and 2-4 indicates that considerable response latitude is available to operators. Several authors in the area of system simulation (e.g., Law & Kelton, 1982) suggest that, in the development of a logical model of a real-world system, it is a good idea to begin with a simple model that can later be made more sophisticated. The initial model should contain only enough detail to meet the basic objectives of the model construction effort.

With this caveat in mind, a decision was made early in the project to restrict the set of operator actions enabled on the POSSUM. The first area of restriction involves the task "Hook Tracks." Current PATRIOT doctrine specifies that the Sequential Hook mode is usually appropriate in situations not involving a priority engagement alert (PEA). Hence, the initial version of the POSSUM is structured to enable only two hook modes: Sequential Hook and Automatic Hook (i.e., the mode used under a PEA condition). The remaining hook modes are not enabled in the model at the present time.

A second area of response restriction on the operator model involves limiting responses to a sequence judged a priori to be optimum. As in the case of the hook modes, the model is to be expanded later to enable less-than-optimal response sequences.

Working within the restrictions noted above, the development of the logical basis for the POSSUM was initiated by specifying optimum operator task flow. Following a suggestion provided by Harris (1981), optimum operator task flow was depicted in a series of directed graphs, or di-graphs. Initially, a separate di-graph was developed for each of the operator task segments under consideration. For example, the di-graph for the operator task "Alert Responding" is presented as Figure 3-2. After separate di-graphs were developed for each task segment, the individual di-graphs were integrated to form a single di-graph representing all operator actions to be modeled. Figure 3-3 presents the final version of the integrated task flow di-graph.

As action begins, the simulated operator is in the Steady State (Monitoring/Review). System cues that result in a transition from Steady State to another system state, in order of their precedence, are given as follows:

System Cue	State Transition
1. Alert message line (AML) contains a priority message.	Alert Process
2. AML contains a blinking "Full" or "More" modifier.	Alert Process
3. Track on situation display circumscribed by broken hexagon.	TBEQ Process
4. Track appears on TBEQ.	TBEQ Process
5. AML contains non-priority message.	Alert Process

If a transition to the Alert Process occurs, the operator remains in that state until all priority messages are cleared and the modifiers "Full" or "More" are removed from the AML. If the operator is in the Alert Process, but has not cleared all non-priority messages, the following system cues preempt the process and result in a transition to the TBEQ Process (through Steady State):

1. Track on TBEQ.
2. Broken hexagon displayed with track.
3. Track number in TBEQ blinking.

If no preemptions occur, the operator clears all messages (priority and non-priority) and then transitions back to Steady State.

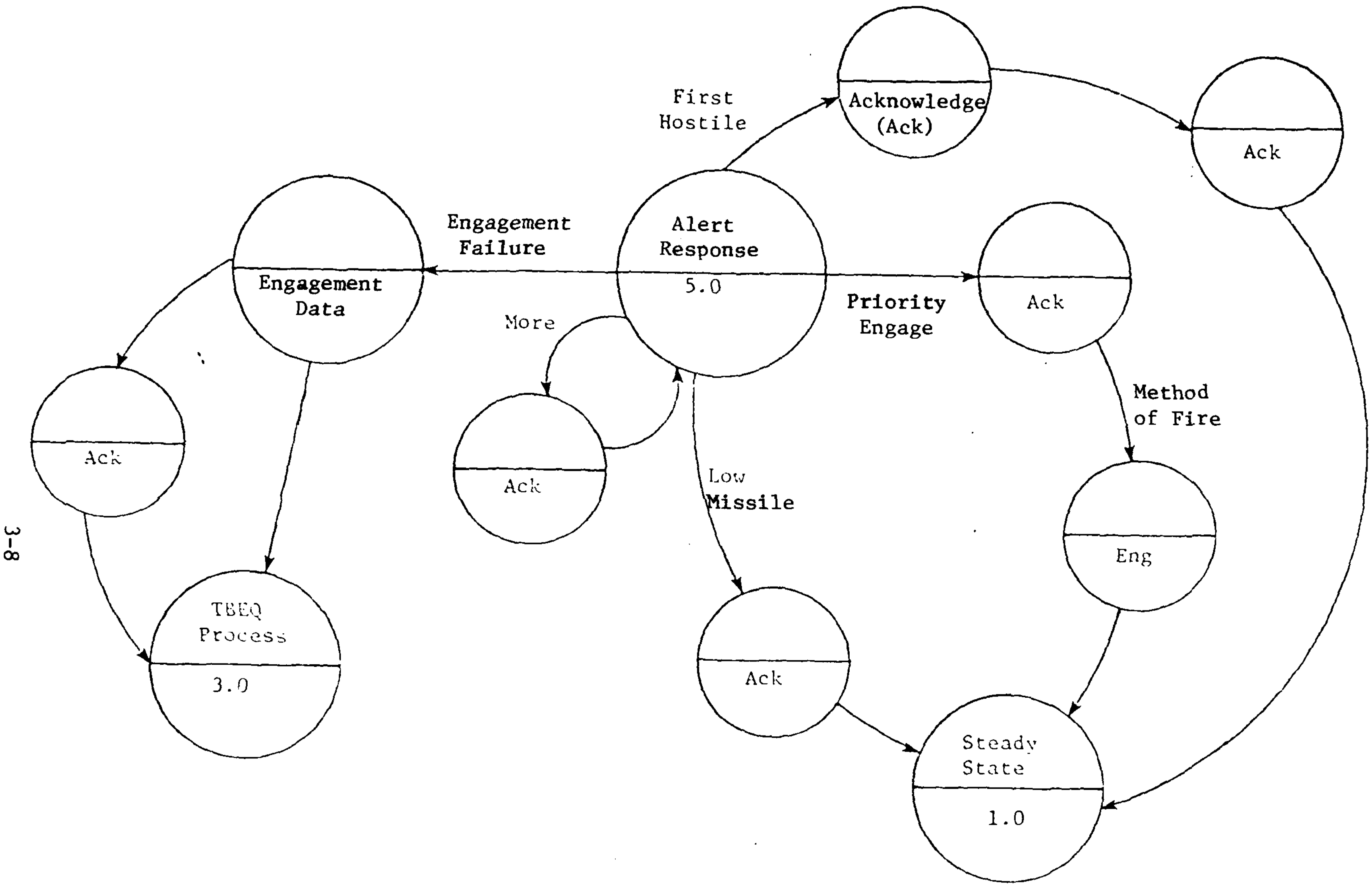

Figure 3-2. Directed Grapn for the Operator Task
Alert Responding

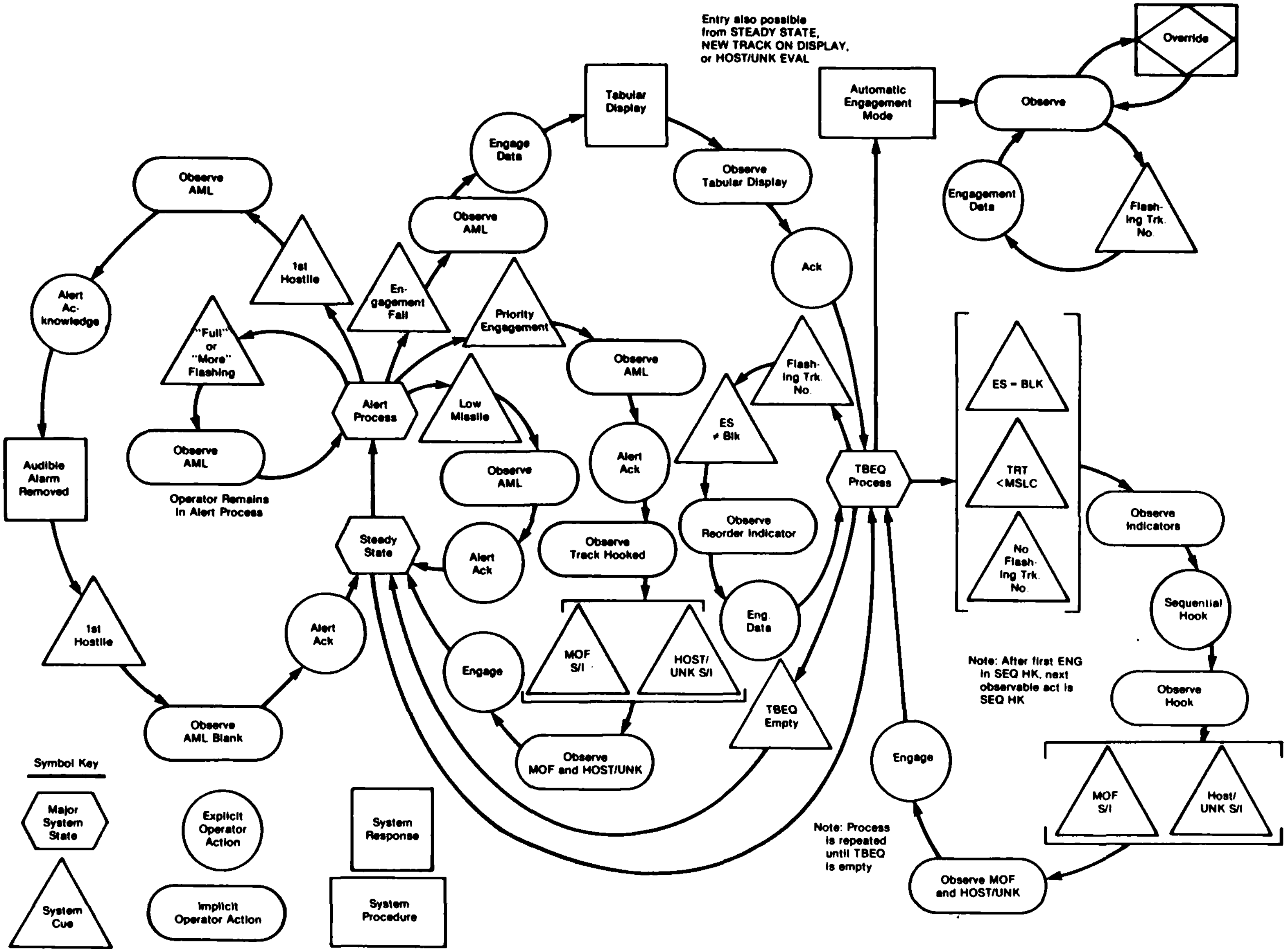

Figure 3-3. PATRIOT Engagement Control Station (ECS) Operator Task Flow (from Howard, C. W., Hawley, J., and Harris, D., in preparation)

Once in the TBEQ Process, the operator carries out the indicated sequence of actions until all tracks have been processed. The TBEQ Process is preempted in favor of the Alert Process by the occurrence of the following system cues:

1. Priority message displayed on AML.

2. AML displayed is not priority, but the modifiers "Full" or "More" begin blinking.

If the operator has completed the TBEQ Process and both the Alert Process and Monitor/Review Process (i.e., Steady State) require a transition, the Alert Process takes precedence. If no conflict is apparent, the operator transitions back to Steady State.

The operator response sequence depicted in Figure 3-3 and described above represents the logical basis for the initial version of the POSSUM. It should be noted again that the logical basis for the initial version of the POSSUM is deterministic. Simulated operator actions explicitly follow the sequence outlined on the di-graph; no operator decision processes are modeled. After a preliminary version of the POSSUM using the optimum sequence of operator actions is developed and evaluated, the logical basis of the model will be expanded to include a range of alternative response sequences. At that point, the POSSUM will become what might be termed a quasi-probabilistic model of a PATRIOT ECS console operator. The term quasi-probabilistic denotes a situation in which probabilistic response choices are made, but from a restricted subset of the universe of potential alternatives.

Given the basic structure for a discrete event simulation and the logical basis presented in Figure 3-3, the operational POSSUM functions as depicted in Figure 3-4. Simulated operator actions are prompted from one of two sources: the system (system cues) or the last action taken by the operator. Provision is also made to queue actions via the action queue in the event the simulated operator becomes overloaded. Each system cue or operator response is followed by a set of lawful following responses. Sets of lawful following responses are identified by the Potential Action Designator (PAD). Designated lawful responses are output as the Potential Action List (PAL). One simulated response is then selected from the PAL in a monte-carlo fashion, taking into account operator characteristics (i.e., psychological and psychomotor profile) and the current situational context (e.g., operator stress and fatigue). In the present version of the POSSUM, only one following response is permitted for each system cue or operator response. Hence, the PAD and PAL are not functional at the current time. Simulated operator responses are uniquely determined from the logical structure described in Figure 3-3.

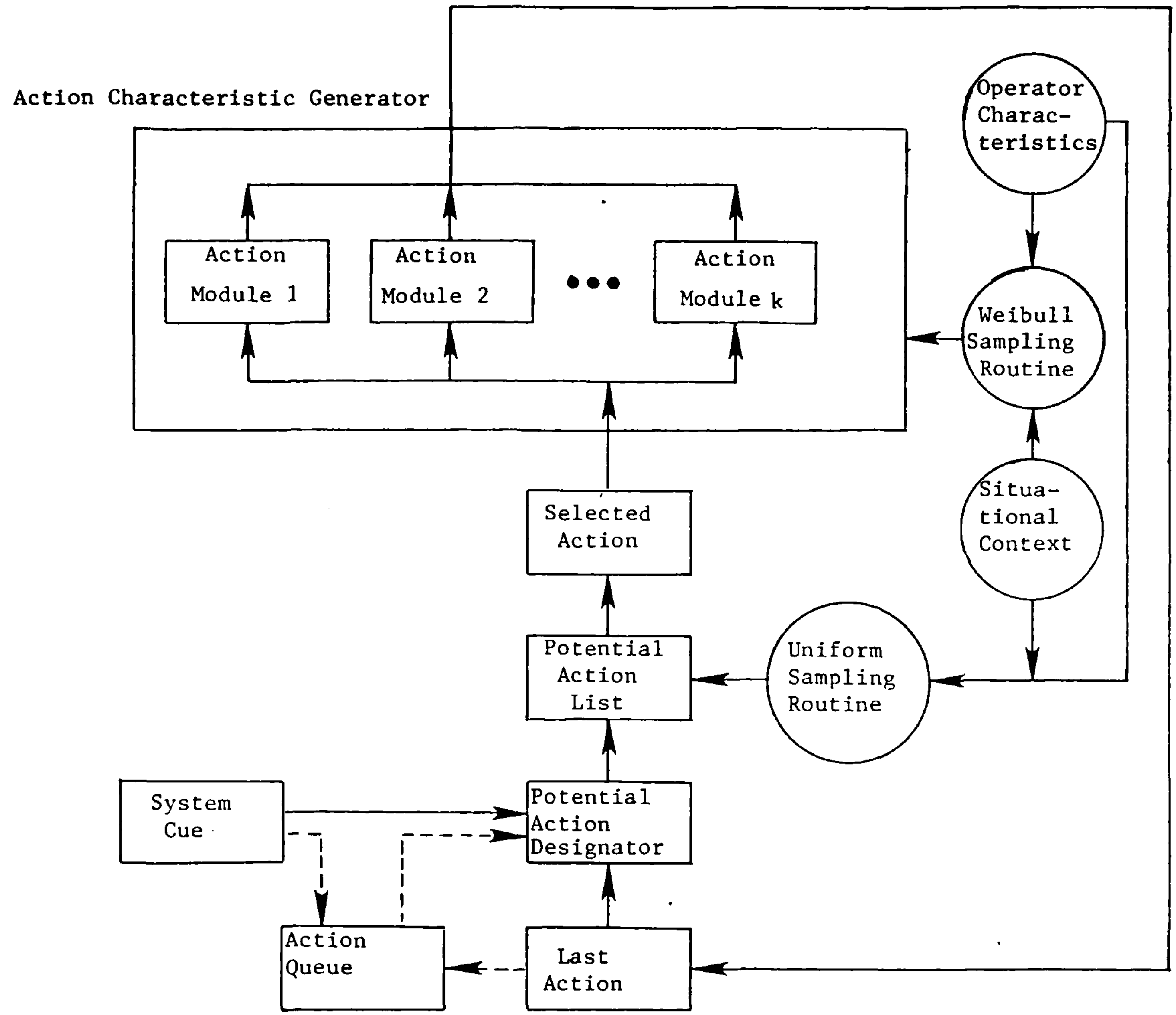

Figure 3-4. POSSUM Concept of Operation

Since the POSSUM functions basically as a discrete event simulation, event times represent a key aspect of the simulation process. Event times in the POSSUM are produced by the Action Characteristic Generator (ACG). The ACG contains an Action Module for each discrete operator action to be modeled. Simulated operator response, or delay, times are obtained by sampling from an appropriate probability distribution. In the POSSUM, a three-parameter Weibull sampling routine is used to simulate the times associated with all operator actions. Operator- and situation-specific response time distributions are determined by selecting Weibull parameters (i.e., location, shape, and scale) to reflect both operator characteristics and current aspects of the engagement environment. More complete documentation on the operational POSSUM is presented in Optimizing Operator Performance on Advanced Training Simulators: Program Documentation.

The POSSUM is intended to simulate purposive, goal-directed operator behavior. It represents the actions of a well-trained, highly motivated operator acting in an optimum fashion. Although the POSSUM does not reproduce exactly all actions of real-world ECS console operators (it is doubtful that any simulation model could do this), it is potentially useful in that a number of valuable inferences concerning the performance potential of the PATRIOT human-machine system can be gained through its application. For example, since the POSSUM is based on an a priori, optimum sequence of operator actions, upper performance limits of the PATRIOT system with a human in the control loop can be determined. The effects on system performance of alternative response contingencies (i.e., operating procedures) or of alternative human-machine function allocation schemes also can be explored by altering the logical structure of the operator model.

Model Application

The initial research plan called for the POSSUM to be integrated with the simulation software on the TOS/T. The POSSUM was to be resident on the TOS/T, rather than developed as a stand-alone procedure, because of the difficulties involved in creating a suitable engagement environment apart from the TOS/T. It was judged that the real-time limitations of having the POSSUM resident on the TOS/T were more than offset by the potential problems of trying to re-create the PATRIOT engagement environment on another computer.

As noted in section 2, repeated problems in scoring test operator performance protocols were encountered. These problems also impacted upon the development and implementation of the POSSUM. In hope that TOS/T system problems would be resolved before the end of the project, a decision was made to develop a preliminary, or test, version of the POSSUM using a separate computer [i.e., on ARI's (Ft. Bliss) Hewlett-Packard (HP) 1000].

Accordingly, a test version of the POSSUM was developed and implemented on the HP (denoted the HP version). This model was debugged and actually exercised. However, given the limited engagement environment available using HP, the resulting simulated operator data was too limited in scope to permit a fair evaluation of the model's utility.

An additional problem limiting the utility of the HP POSSUM is a lack of valid data with which to parameterize the model. Parameterization refers to a process of defining model parameters (e.g., mean response times, response time distributions, choice response probabilities, etc.) on the basis of the observed behavior of the real system (Shannon, 1975). In the case of the POSSUM, model parameters were to be estimated from the test operator performance protocols, which, as noted earlier, were not scored.

The upshot of the preceding discussion is that several of the project's objectives were not completely met. Specifically, objective three, the development of an operator optimization model, was only partially met. The operator simulation model was developed, but not suitably implemented. Objectives four and five were not addressed at all. During the course of the project, however, a number of developments occurred that caused the authors to alter their preconceptions of how the latter aspects of objective three (i.e., model parameterization) and objectives four and five should be addressed. Since these developments represent a departure from what was planned initially, and because the material is not documented elsewhere, it is appropriate (as well as informative) that they be included as part of this report. Accordingly, procedures for parameterizing the POSSUM and for validating the operator model are discussed in the next portions of the report.

Model Parameterization

In order for a simulation model to function, model parameters such as response time distributions, state-transition probabilities, and the like must be defined either a priori or on the basis of the behavior of the real system (i.e., the entity being modeled). Model parameterization refers to the process of defining these characteristics of system performance. In the current effort, model parameterization consists of two aspects, listed as follows:

1. Simulated operator response characterization
2. The consideration of situational factors that moderate operator performance

Implied in the full definition of the POSSUM is a third aspect of parameterization: the treatment of state-transition probabilities. However, since

the initial version of the POSSUM is deterministic, the consideration of state-transition probabilities is not relevant to its development at this time. Their consideration is thus relegated to such time as a probabilistic version of the POSSUM is developed. Each of the remaining aspects of model parameterization is discussed in the following paragraphs.

Simulated Operator Response Characterization

After a simulated operator action has been determined (reference Figure 3-4), the next aspect of the operator simulation involves specifying characteristics of the selected response. For most switch actions and key presses, this step consists of specifying the time to complete the action. In the POSSUM, simulated operator response times are derived by sampling from an appropriate theoretical probability distribution. A problem that often arises in system simulation concerns first identifying an appropriate theoretical probability distribution for describing empirical phenomena (e.g., normal, lognormal, beta, gamma, etc.), and then selecting the correct parameters for that distribution (e.g., mean, standard deviation, etc.). Several authors (e.g., Shannon, 1975; Law & Kelton, 1982) have described procedures for identifying appropriate theoretical probability distributions from empirical data. These procedures are, however, typically quite cumbersome and time consuming in application. In addition, the procedures often prescribe a separate probability distribution for each event being modeled, thus increasing the complexity of the simulation software.

A potential shortcut to the use of separate probability distributions for different events is to use the three-parameter Weibull distribution to model all simulation events. The Weibull is a highly flexible distribution characterized by three parameters: location (a), shape (b), and scale (c). By appropriately selecting the three parameters, a variety of shapes for the density function are obtained. Mills and Hatfield (1974) report that the Weibull distribution, used in this manner, is consistently accurate in fitting observed task performance completion time distributions in a sequential task performance situation. In view of this positive evidence, a decision was made to employ the three-parameter Weibull to characterize simulated operator responses in the POSSUM.

The decision to use the three-parameter Weibull to simulate all operator responses eliminates one of the major problems of model parameterization: the choice of an appropriate distribution family. However, the parameter estimation problem remains to be addressed. In the POSSUM, Weibull parameters are obtained using a method described in Zanakis (1979). Following Zanakis' description, a derivative-free, pattern search nonlinear optimization procedure for obtaining maximum likelihood estimates (MLEs) of Weibull parameters was developed. Basically, the Weibull MLE program functions as follows:

1. A set of empirical response duration times--t_1, t_2, ..., t_n-- is read and sorted into descending order.

2. Initial estimates for the location, scale, and shape parameters are computed. The initial estimates are denoted $\hat{a}_o$, $\hat{b}_o$, and $\hat{c}_o$, respectively.

3. Bounds for $\hat{a}$, $\hat{b}$, and $\hat{c}$ are established.

4. Using the Weibull MLE program, values for $\hat{a}$, $\hat{b}$, and $\hat{c}$ that maximize the Weibull log-likelihood function, $L(\hat{\theta})$, subject to the constraints established in (3), are determined.

After MLEs for _a_, _b_, and _c_ have been obtained, the MLE program presents graphs of the cumulative distribution functions for the empirical response times and for the best-fitting Weibull, superimposed upon each other. A Kolmogorov-Smirnov procedure is then used to evaluate statistically the fit of the best Weibull to the empirical data. This latter step is taken to safeguard against problems of non-convergence, local maxima, and so forth that are often associated with the application of iterative optimization procedures like the Weibull MLE program. Documentation for the Weibull MLE program is provided in _Optimizing Operator Performance on Advanced Training Simulators: Program Documentation_.

The Treatment of Situational Factors That Moderate Operator Performance

The application of the Weibull MLE procedure to empirical test operator data provides the parameters necessary to characterize simulated operator responses in the POSSUM. There is, however, another aspect of model parameterization that should be considered prior to describing the actual response generation process. This additional aspect of parameterization concerns the treatment of situational factors that are expected to moderate operator performance. Although numerous potential moderator variables have been identified and discussed in the human performance literature, only two, denoted herein as stress and fatigue, are explicitly considered in the POSSUM. In this portion of the report, these constructs are discussed and operationally defined; their use in the generation of simulated operator response times is described in the next portion.

Swain and Guttman (1980) define stress as the human response to a stressor. A stressor is defined as any external or internal force that causes bodily or mental tension. Following this definition, stressors are separated into two classes: physiological and psychological. Examples of each class of stressor are listed as follows:

Physiological	Psychological
Fatigue	Task Speed
Discomfort	Distractions
Constriction of Movement	Monotony
High Temperature	Emergency Situations

In terms of human response to stress, Edward and Lees (1973) list five typical reactions:

1. Queueing - delaying some responses during overload.
2. Omission - ignoring information or actions that are considered relatively unimportant.
3. Gross discrimination - responding to gross aspects of stimuli but ignoring finer aspects.
4. Errors - processing information incorrectly.
5. Escape from task - physical or mental withdrawal.

All of these reactions can serve to moderate console operator performance, thus it was judged important to adjust simulated operator actions in the POSSUM to reflect real-world human operator reactions to such situational conditions.

Following a review of the literature on human response to stress, as broadly defined by Swain and Guttman and primarily within the context of human operator modeling (e.g., Conrad, 1956; Siegel & Wolf, 1969; Edward & Lees, 1973; McCormick, 1976; Pew, Barron, Feehrer & Miller, 1977; Hixson & Grant, 1980; Swain & Guttman, 1980), a decision was made to treat the two stress categories separately.

Considering first psychological stress (or "stress" as the term was used earlier), Swain and Guttman (1980) note that objective data on performance under stress are spotty. No comprehensive treatment of the effects of stress on performance is presented in the literature. However, Conrad (1956) states that performance under stress is typically a linear function of the product of load and speed. In this context, load is defined as the variety of stimuli (type and number) to which a receiver must attend; speed is defined as the number of stimuli occurring per unit of time (McCormick, 1976). Note that this characterization of stress is the same as that used in the definition of instantaneous situational difficulty in section 2 [expression (2-6)]. Accordingly, operator stress at time t is operationally defined to be the value of the function $\Phi(t)$.

The consideration of physiological stress (or "fatigue", to use an earlier term) presents a somewhat easier problem than psychological stress. All of the aspects of physiological stress listed earlier increase as a function of the time the operator is in a physically discomforting environment. Hence, it is not unreasonable to operationally define the level of physiological stress at time t in terms of the total length of time the operator has been in an operational environment.

The Response Generation Process

As noted earlier in the discussion of the POSSUM, it is desired that simulated operator responses reflect both operator characteristics and the situational context (i.e., stress and fatigue). Since simulated operator responses are to be characterized through the selection of Weibull parameters, this desideratum implies that the Weibull parameters must reflect operator and situational factors. That is, a means for selecting Weibull parameters on the basis of operator characteristics and situational factors must be devised.

One means of obtaining Weibull parameters that are sensitive to operator and situational characteristics is to use the situational factors as independent variables in a regression model with the Weibull parameters as criteria; that is,

$$(\hat{a},\hat{b},\hat{c}) = f(\hat{P},\Phi,F). \tag{3-2}$$

In (3-2), $\hat{a},\hat{b}$, and $\hat{c}$ represent Weibull parameters;

$\hat{P}$ represents expected operator performance;

Φ represents operator stress;

F represents operator fatigue;

and $f(\cdot)$ represents a polynomial function relating $\hat{P}$, Φ, and F to $\hat{a},\hat{b}$, and $\hat{c}$.

Note that in (3-2) operator characteristics are represented by the term $\hat{P}$. In application, $\hat{P}$ is estimated from operator psychological and psychomotor characteristics; that is, $\hat{P} = g(\underline{\Psi})$, where $\underline{\Psi}$ is a vector of psychological/ psychomotor characteristics.

The only problem in implementing the procedure described above is that values for the dependent variable set--$\hat{a}$, b, and $\hat{c}$--do not exist a priori. These values are determined from an array of empirical task completion times using the Weibull MLE program. To provide reasonable parameter estimates, the times in the array should be associated with similar, or nearly similar, values of $\hat{P}$, Φ, and F. As a result of this constraint, standard regression procedures for estimating the parameters of (3-2) are not applicable.

A solution to this apparent dilemma, suggested by Law (1981), is to estimate Weibull parameters using what is referred to as a "segmentation" approach. Using the segmentation approach, the independent variables $\hat{P}$, Φ, and F are each categorized resulting in a three-way contingency table, as depicted in Figure 3-5. For simplicity's sake only three levels of each factor are shown in Figure 3-5. After selecting appropriate category boundaries for each factor, empirical task completion times are sorted into the cells of the matrix. The MLE program is then exercised on the arrays of times within the cells of the matrix. The result is a set of Weibull parameters for each cell (as shown in Figure 3-5). Later, judgmental or statistical procedures are used to determine whether all of the separate factors, or levels within factors, are necessary; that is, to determine whether sufficient separation exists between Weibull parameters to warrant retaining all factors or factor levels. It should be noted that this process is repeated for each operator response type to be modeled.

Law's procedure provides Weibull parameter estimates that are sensitive both to operator characteristics and to situational variables. The method is employed quite simply in the POSSUM via a table look-up procedure. For a specified class of operators, $\hat{P}$ is computed in advance on the basis of psychological and psychomotor characteristics (i.e., $\underline{\Psi}$). When the POSSUM requires a simulated operator response, current values for Φ and F are computed. These three values determine a specific cell in a response-type matrix containing Weibull parameters. The selected parameters are then input to a Weibull random number generator resulting in a simulated response completion time.

Having now described the operation of the POSSUM, which is to be used as a partial surrogate for experimentation with actual PATRIOT console operators, the final step before applying the model is validation. Prior to employing the model, it is necessary to ascertain the extent to which the operator simulation model is an accurate representation of the behavior of actual operators. The next portion of the report outlines procedures for the conduct of validation studies on the POSSUM.

Model Validation

Validation refers to the process of determining whether a simulation model is an accurate representation of the actual system being studied. In most simulation situations, there is no definitive test for model validity. This results from the fact that a simulation model, regardless of how complex, is usually only an approximation to the real system. As a result, a series of evaluations directed at validation issues is conducted throughout the model development process. The ultimate objective of the model evaluation process is to build user confidence that inferences derived from the application of the simulation model are correct (Shannon, 1975).

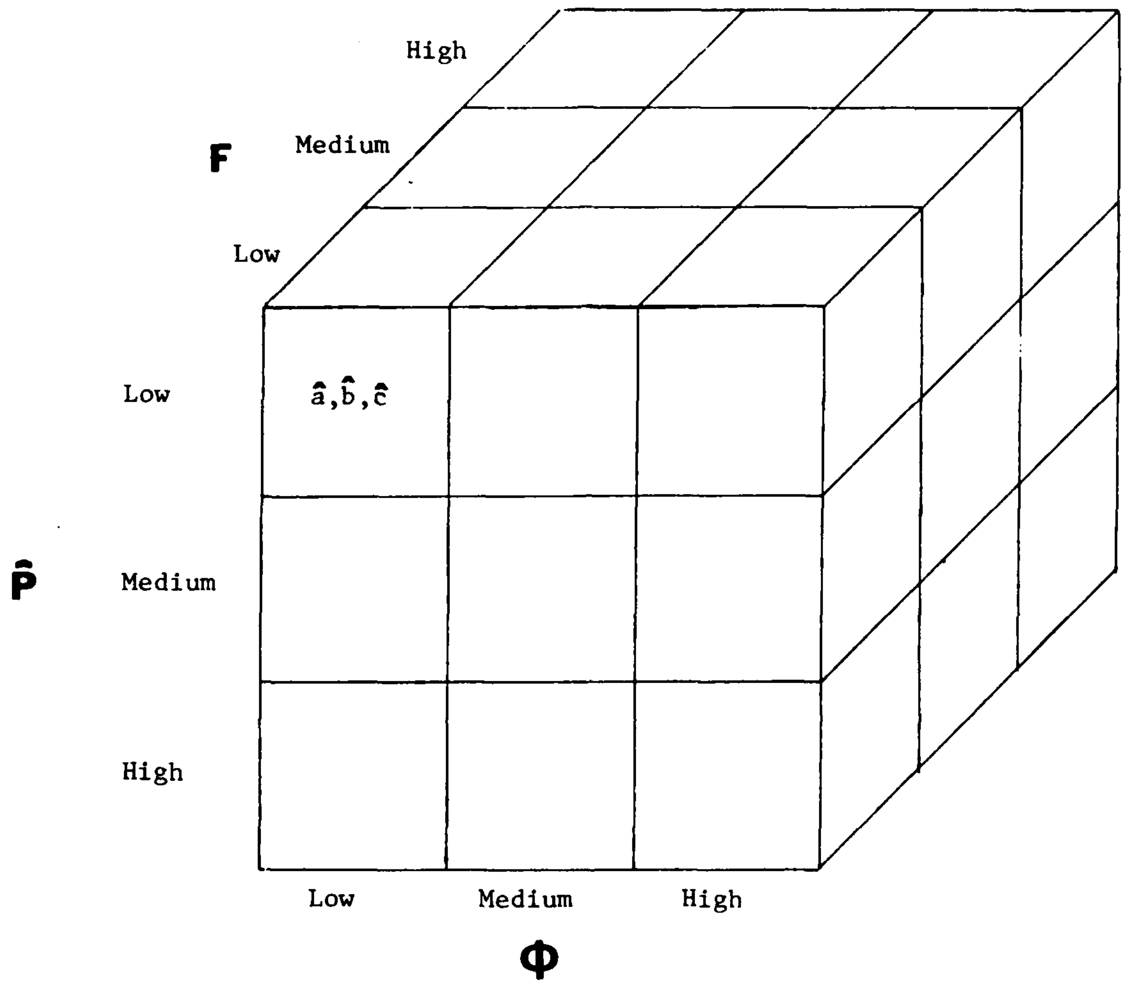

Figure 3-5. POSSUM Segmentation Scheme

Simulation model validation is typically characterized by two phases (Law & Kelton, 1982):

1. Verification
2. Validation, per se

Verification addresses the issue of whether the simulation model performs as intended. The objective of the verification step is to eliminate logical errors in the model's structure, its mathematical algorithms, or the corresponding computer programs.

Actual model validation typically addresses three additional issues (Fishman & Kiviat, 1967; Naylor & Finger, 1967; Hermann, 1967; Rivett, 1980; Law & Kelton, 1982):

1. Face validity
2. The validity of the model's underlying assumptions
3. Predictive validity

Face validity concerns the reasonableness of the model's output; that is, the extent to which the simulation model produces results that are similar to the output of the real system. Turing (1959) has proposed a test for reasonableness. This test consists of locating persons who are familiar with the behavior of the real system (i.e., actual operators) and asking them to compare simulation results with output from the real system. If the panel of experts cannot differentiate real system output from simulated output, the model is judged to have face validity.

Validation of the model's underlying assumptions concerns verifying model assumptions through experimental testing. This step usually addresses two sub-issues (Hermann, 1967):

1. Does the simulation produce low variation in output when replicated with all external inputs held constant? If the model has high variability of output due to internal processes, then it is doubtful that the relationships assumed in the model accurately reflect the real world.

2. Do relationships between variables in the simulation correspond to those of the real world? For example, is the simulation model as sensitive in its reaction to changes in its parameters as the real world appears to be?

In a sense, step two is a quantitative extension of step one.

The final aspect of validation is using the simulation model to predict real system behavior. In the majority of modeling efforts, predictive

validation constitutes the ultimate test of model validity. Two aspects of predictive validation are usually employed: historical or retrospective validation and forecasting or prospective validation. Retrospective validity concerns the model's ability to replicate statistically previous behaviors of the real system. Prospective validation concerns the capabilities of the model in accurately predicting the behavior of the real system in new situations.

In the current effort, the intention is to conform to this suggested validation process in evaluating the POSSUM. First, the POSSUM will be used to produce analogs of test console operator performance protocols. The results from these simulation runs will then be compared with actual operator protocols using a panel of experts selected from the Air Defense community at Fort Bliss. This comparison will constitute a Turing test of model face validity.

Following a sufficient number of replications, it will be possible to evaluate the POSSUM's internal consistency and sensitivity (validation step 2). Standard statistical procedures for characterizing and evaluating simulation output will be used to this end [see Fishman (1978) or Law & Kelton (1982) for a detailed discussion of the statistical treatment of simulation output].

After the POSSUM is subjected to face and empirical validation checks, the final step in model evaluation is predictive validation. In this regard, number of statistical tests for comparing simulation model output with the behavior of a corresponding real system have been proposed (e.g., Shannon, 1975; Fishman, 1978). Such statistical comparisons are not as straightforward as they might appear, however. Since the output of nearly all simulation models (as well as that of their real-world counterparts) is nonstationary and autocorrelated (i.e., is the result of a nonstationary, autocorrelated stochastic process), the use of classical statistical procedures typically is not appropriate. However, even if a direct statistical comparison were appropriate, it is doubtful whether testing for differences between a simulation model's behavior and the behavior of the real system is reasonable. Since a simulation model is usually only an approximation of the real system, a null hypothesis that the model's performance and the real system's performance are identical will nearly always be rejected. Law and Kelton (1982) suggest that a more appropriate concern is whether observed differences between the real system and a simulation model will affect conclusions, vis-à-vis the real system, derived through the application of the simulation model. As an approach to predictive validation, Law and Kelton's position reflects the view that model validation should primarily concern the worth of the insights gained through the use of a simulation model, rather than a demonstration of the model's ability to replicate exactly the behavior of the real system.

In keeping with the notion discussed above, output data from the POSSUM and from actual operators will be compared using a series of control charts. A representative control chart is shown as Figure 3-6. Mean test operator performance scores (for the SPM and each of the MPMs) from each scenario are displayed, along with their respective confidence interval bands. For ease of interpretation, scenarios are ordered along the X-axis according to their difficulty levels. Also shown on the control chart are mean POSSUM performance scores (i.e., the results of a series of POSSUM replications) and a profile representing the performance of the PATRIOT system in automatic mode. If desired, confidence bands about mean POSSUM performance can also be displayed. Given the logical structure of the POSSUM, its mean performance profile should fall somewhere between the profile for automatic and that of actual operators. The control chart approach to predictive validation permits a rapid visual evaluation of POSSUM results. Statistical comparisons via the superimposed confidence intervals also are facilitated. Obvious flaws in the POSSUM's performance should be readily identifiable from the control charts.

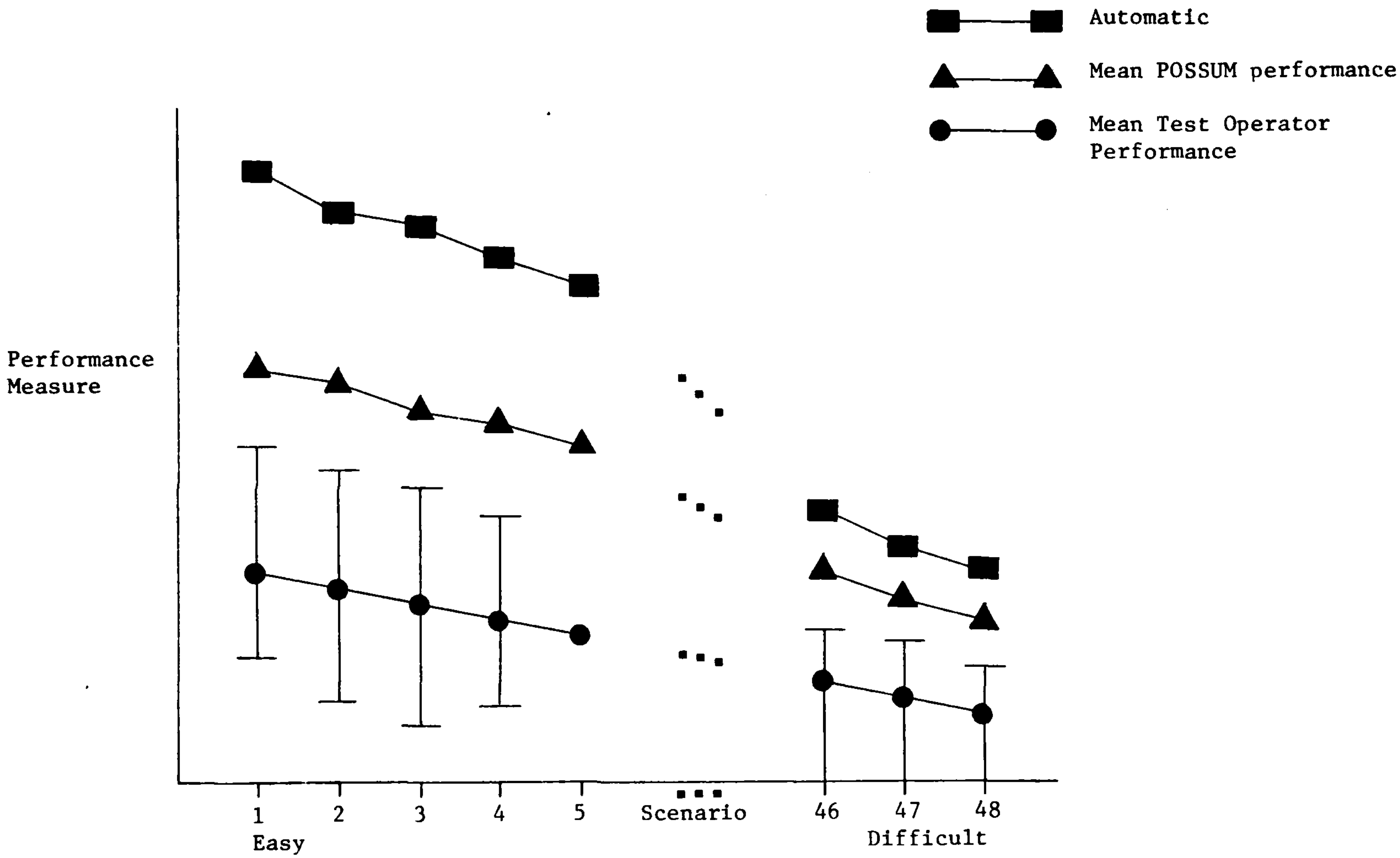

Figure 3-6. Exemplary POSSUM Control Chart

4. DEVELOPMENTAL ISSUES

This report presents results from the first year of a research effort concerned with the general topic of human-machine integration in automated systems. The specific focus of the effort is on the PATRIOT Air Defense missile system, but the methodology holds promise for application in other computer-aided human-machine systems, such as air traffic control or anti-submarine warfare. As noted several times during the report, the primary thrust of the effort is the development of a vehicle for enhancing overall system performance through a systematic consideration of performance shaping factors such as: (1) operator selection, (2) operator training, (3) human-machine integration, and (4) system employment.

Prior to considering the performance enhancement problem, the first requirement in the study concerned the quantification of human operator performance. Following a framework for performance assessment in human-machine systems adapted from Meister (1976), a hierarchical network of operator performance measures was developed. The performance assessment scheme addresses the system, mission, and personnel levels of functioning in the PATRIOT ECS environment. However, at this preliminary stage of development, only a single ECS console operator is considered. The operator's interaction with other components of the PATRIOT system is not addressed.

The PATRIOT ECS console operator performance evaluation scheme also considers a concomitant aspect of performance assessment: situational difficulty as a moderator of performance. In this regard, a situational difficulty index to be used in adjusting operator scores to reflect scenario task demands was developed. Procedures for computing and applying the SDI are presented and discussed.

In accord with project plans, the operator performance assessment capability was implemented on the TOS/T, an environmental, full-task simulator for the PATRIOT system. Plans called for exemplary operator scores to be obtained by using the evaluation capability to score the performance protocols of a series of test console operators. The intent of this exercise was to demonstrate the utility of the performance assessment capability and to provide data required in the development of the operator performance optimization model. However, technical problems with the TOS/T prevented the completion of most planned scoring activities.

The second major project activity concerned the development of an operator performance optimization model. This model, which is to relate human-machine performance to various classes of performance shaping factors, is to be used to enhance total human-machine performance prior to fielding the PATRIOT system. Due to the magnitude of the performance enhancement problem, the development of the performance optimization model

was approached through the construction of a computer simulation model of a PATRIOT ECS console operator. As noted in section 3, this simulation model (denoted the POSSUM) is to be used as a partial surrogate for experimentation with actual operators.

Following the above perspective, a preliminary version of an operator simulation model was developed. This initial version of the model is restricted in its function in that it enables only a subset of operator actions and then in a judged optimum fashion. As in the case of the scoring capability, original project plans called for implementing the POSSUM on the TOS/T, then validating the model and using it to provide data necessary for the development of a regression-type performance prediction/optimization model. However, the same problems that prevented the complete installation of the scoring capability on the TOS/T also delayed the implementation of the POSSUM. As a result, the only POSSUM results currently available were provided by a test version of the preliminary model. Problems in creating a realistic engagement environment apart from the TOS/T and in obtaining data with which to properly parameterize the model limit the utility of these initial POSSUM results. A cursory face validation appraisal of the initial POSSUM results (by the project staff) indicates, however, that the model functions as intended and provides reasonable output, considering the aforementioned problems.

Discussion

Although the technical objectives of the project were not completely met, considerable groundwork in the areas of operator performance assessment and modeling was laid. In the area of performance evaluation, probably the most important contributions are: (1) a clarification of Meister's framework for human-machine performance evaluation, and (2) an application of this framework using the PATRIOT Air Defense missile system as an exemplar. In this exemplary application, it was demonstrated that operator performance can be quantified at a variety of levels and that the resulting data are reasonable.

A second major contribution of the current project concerns the treatment of situational difficulty as a modifier of operator performance. Previous efforts (e.g., Sheldon & Zagorski, 1965) have recognized the necessity of adjusting raw operator performance indices to reflect situational difficulty, but satisfactory methods for treating the problem were not forthcoming. Quite surprisingly, though, a number of recent authors in the area of human-machine performance assessment (e.g., Callan, Kelley, & Nicotra, 1978; Connelly, 1981; Obermayer & Vreuls, 1974) have not addressed the issue of situational difficulty at all. Admittedly, the treatment of situational difficulty as described herein is not definitive, but the approach holds promise for the future.

The most notable shortcomings of the current effort involve the failure to implement completely either the scoring capability or the operator simulation model on the TOS/T. These failures precluded obtaining the data necessary to demonstrate the utility of the operator evaluation scheme or the validity of the POSSUM. As a result, both of these activities remain to be completed. Upon a review of the TOS/T implementation problems encountered in the current project, several observations are in order. The first observation concerns the long "learning curve" for the project staff. Integrating additional code into a simulator as complex as the TOS/T is not a trivial undertaking. A competent, stable, and dedicated programming staff is required to carry out software modifications of the type attempted. As an addendum to this observation, the time involved in making required software modifications should not be underestimated. Even with an able programming staff and given only minimal hardware problems, a considerable expenditure of time and resources is required.

Future Directions

The activities reviewed in this report concern a single ECS console operator and are limited to a subset of operator tasks. As noted in section 1, a single ECS operator represents only one component of the total PATRIOT human-machine system. Thus, in order to be made directly relevant to the real-world, the work reported herein would have to be extended in several directions. Perhaps the most significant extension of the current work involves a consideration of multiple operators and multiple operator stations. In the case of performance assessment, such an extension would include the definition of criteria for individual operator positions throughout the PATRIOT battalion, as well as the development of performance measures for the ECS team, for the ICC team, and for the battalion operating as a unit. It is anticipated that such an augmented performance evaluation scheme would be considerably more complex than the current capability. The increase in complexity would result from the treatment of aspects of C^3 that are not considered in the performance measures for a single operator.

A second issue to be addressed as part of the development of an expanded operator performance assessment capability is situational difficulty. First of all, there are several problems in the definition and use of the current SDI. For example, the current SDI has been criticized as not being sensitive to the position of hostile and unknown tracks vis-à-vis defended assets (Harris, 1981). A second problem with the current SDI is its computation. As noted in section 2, operators interact with the engagement environment dynamically either to increase or to decrease situational task demands. What is desired in the SDI is an a priori index that can be derived independent of operator or system performance. The current SDI does not completely meet this expectation.

For the reasons noted above and for others involving: (1) the theoretical basis of the SDI, and (2) its generalizability to team and multi-team operations (e.g., the current SDI does not reflect operator loading due to voice communications or other aspects of command and coordination), the issue of situational difficulty should be examined in greater detail. A valid SDI is important in reporting normative operator performance. The index could also provide the basis for a rational scenario design capability, which would be extremely useful as part of a progressive operator training/evaluation program.

A third topic to be addressed in a future effort is the continued development of the POSSUM. First of all, the current version of the model has not been parameterized or subjected to any validation studies. Hence, the first task in a renewed model development effort would be the completion of work left outstanding from the current project; that is to parameterize and validate the POSSUM.

Having completed the work outstanding from the current effort, an obvious next step in the development of the POSSUM is to expand its logical basis and thus to provide the model with the capability of simulating a broader range of operator behavior. Following this step, the model could then be further expanded to provide an operator team and multiple operator team simulation capability. An expanded performance modeling capability of this type would permit training designers and combat developers to address relevant system development issues (e.g., selection, training, etc.) in a more complete fashion than permitted under the current version of the POSSUM.

To illustrate the potential utility of the performance assessment and modeling capabilities described in this report, consider the training/evaluation process illustrated in Figure 4-1. To begin the process, each scenario selected for use in training or evaluation is run using the system in automatic mode. The results of this run are SP and MP scores for that scenario. Situational difficulty is also computed and is used later in providing normative evaluatee scores. All of these results--SP, MP, and SD--are entered into a standards data base, which is made available to trainers and evaluators.

Essentially the same process described above is repeated using the POSSUM. SP, MP, and TP data are obtained for major classes of trainees/evaluatees. These data provide information regarding the relative, or expected, levels of performance for various classes of trainees/evaluatees.

Actual PATRIOT trainees/evaluatees are put through the process in the following manner. First, the operator is evaluated using a scenario for which absolute and relative performance indices are known. In an institutional setting, evaluation is conducted using the Operator Tactics Trainer;

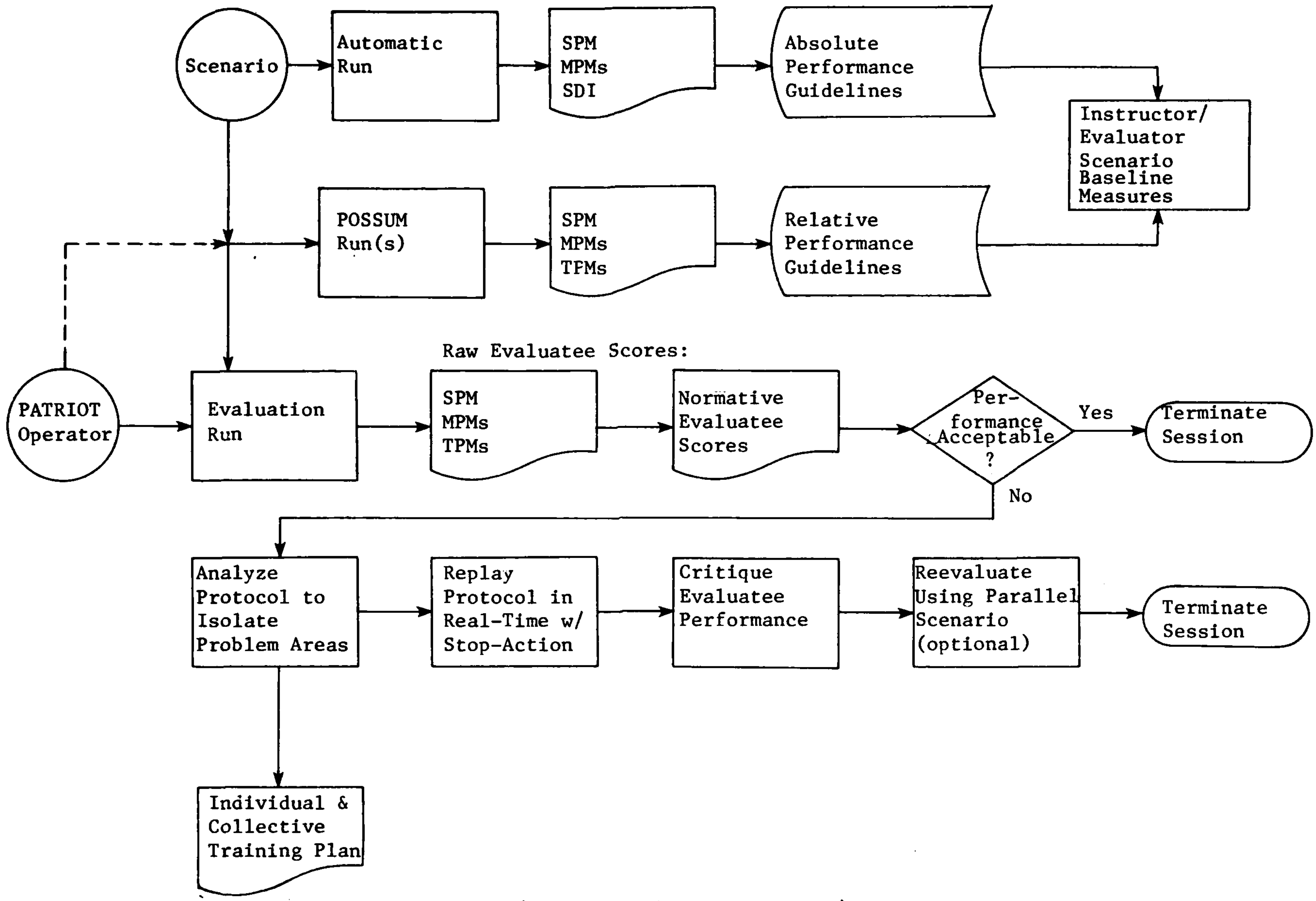

Figure 4-1. PATRIOT Trainee/Operator Embedded Training/Evaluation Process

field personnel are evaluated using the Troop Proficiency Trainers that accompany the PATRIOT system. The results of the evaluation exercise are operator SP, MP, and TP scores. Using the adjustment procedure described in section two, normative SP and MP results are obtained. From these normative results, it is possible to determine whether the operator's performance is satisfactory. If operator performance is satisfactory, the evaluation session is terminated.

For those situations in which the operator's performance is not satisfactory, an additional series of evaluation activities is initiated. The first step in this follow-on evaluation process involves analyzing the operator's performance protocol to identify problem areas. This action is taken to determine the reasons for unacceptable levels of performance at the system- or mission-descriptive levels. The source of this diagnostic information is typically the operator's TP data.

After isolating the sources of the operator's performance difficulties, the next step in the critique process is to replay the protocol for the evaluatee. When the instructor/evaluator spots a specific problem area (i.e., the use of an inappropriate hook mode), the replay is stopped and the evaluatee coached regarding ways to improve his performance. The instructor/evaluator has the option of allowing the evaluatee to resume real-time action at any point in the replay. If this option is exercised, the evaluatee is provided with immediate feedback, via new SP and MP scores, on the effects of changes in his behavioral repertoire. A powerful tool for providing evaluatees with knowledge of results is thus created. If desired, new behavior on the part of the evaluatee is reinforced through re-evaluation using parallel scenarios.

References

Angoff, W.H. Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council on Education, 1971.

Bobillier, P.A., Kahan, B.C., & Probst, A.R. Simulation with GPSS and GPSS V. Englewood Cliffs, N.J.: Prentice-Hall, 1976.

Bock, R.D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.

Callan, J.R., Kelley, R.T., & Nicotra, A. Measuring submarine approach officer performance on the 21A40 trainer: Instrumentation and prelininary results (NPRDC-TR-78-9). San Diego, Calif.: U.S. Navy Personnel Research and Development Center, January 1978.

Connelly, E.M. Development of system performance measures (Technical Report). Vienna, Va.: Performance Measurement Associates, 1981.

Conrad, R. The timing of signals in skill. Journal of Experimental Psychology, 1956, 51, 365-370.

Edwards, E., & Lees, F.P. Man and computer in process control. London: The Institution of Chemical Engineers, 1973.

Emshoff, J.R., & Sisson, R.L. Design and use of computer simulation models. New York: MacMillan, 1970.

Final Report Prototype Qualification Test--Government (PQT-G) of Development Test II (DT-II) of PATRIOT Air Defense System, (and supplemental letter reports). Aberdeen Proving Ground, Md.: U.S. Army Test and Evaluation Command, October 1980.

Fishman, G.S. Principles of discrete event simulation. New York: Wiley, 1978.

Fishman, G.S., & Kiviat, P.J. The analysis of simulation-generated time series. Management Science, 1967, 13, 525-557.

FM 44-15-1 (Test) Operations and Training, PATRIOT, February 1979.

FM 44-15 (Test) Air Defense Artillery Employment, PATRIOT, February 1979.

Gillett, B.E. Introduction to operations research: A computer-oriented algorithmic approach. New York: McGraw-Hill, 1976.

Harris, D.C. Personal Communication, January 23, 1981.

Hermann, C. Validation problems in games and situations. _Behavioral Science_, 1967, _12_, 407-423.

Hixson, A.F., & Grant, F.H. _A SAINT model of the AN-TSQ-73 Missile Minder System: A statistical analysis_. Contract MDA903-79-M-0115. West Lafayette, Ind.: Pritsker & Associates, April 1980.

Hoffer, P.L., & Howard, C.W. _Human factors evaluation of an air defense training simulator--TOS/T_ (Working Paper). Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, September 1979.

Hopkin, V.D. The measurement of the air traffic controller. _Human Factors_, 1980, _22_(5), 547-560.

Hosch, H.A., Starner, J.W., & Howard, C.W. _Development and evaluation of measurement indices of test scenarios for the PATRIOT Air Defense System_. Contract MDA903-79-C-0388. El Paso, Tex.: The University of Texas at El Paso, January 1980.

Howard, C.W. Methodology for evaluating operator performance on tactical operational simulator/trainers. Paper presented at the 20th Annual Conference of the Military Testing Association, October 1978.

Howard, C.W. Psychological profiles and performance characteristics of tactical simulator console operators. Paper presented at the 21st Annual Conference of the Military Testing Association, October 1979.

Howard, C.W. _Criterion based evaluation procedure for PATRIOT console operator performance analysis_ (Working Paper). Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, October 1979.

Howard, C.W. Personal Communication, November 15, 1980.

Jorgensen, C.C., & Strub, M.H. _Analysis of manual threat evaluations and weapons assignment (TEWA) in the AN/TSQ-73 air defense system_ (Technical Report). Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, 1979.

Law, A.M. Personal Communication, January 23, 1981.

Law, A.M., & Kelton, D. _Simulation modeling and analysis_. New York: McGraw-Hill, 1982.

McCormick, E.J. _Human factors in engineering and design_. New York: McGraw-Hill, 1976.

Meister, D. Human factors: Theory and practice. New York: Wiley, 1971.

Meister, D. Behavioral foundations of system development. New York: Wiley, 1976.

Mills, R.G., & Hatfield, S.A. Sequential task performance: Task module relationships, reliabilities, and times. Human Factors, 1974, 16(2), 117-128.

Naylor, T.H., & Finger, J.M. Verification of computer simulation models. Management Science, 1967, 14, 92-101.

Obermayer, R.W., & Vreuls, D. Combat-ready crew performance measurement system, Phase II: Measurement system requirements (USAFHRL-TR-74-108(III)). Wright-Patterson AFB, Ohio: U.S. Air Force Human Resources Laboratory, December 1974.

Optimizing Operator Performance on Advanced Training Simulators: Program Documentation. Valencia, Pa.: Applied Science Associates, Inc., March 1982.

Pew, R.W., Baron, S., Feehrer, C.E., & Miller, D.C. Critical review and analysis of performance models applicable to man-machine systems evaluation. Contract F44620-76-C-0029. Cambridge, Mass.: Bolt, Beranek, and Newman, March 1977.

Rivett, P. Model building for decision analysis. New York: Wiley, 1980.

Rouse, W.B. Systems engineering models of human-machine interaction. New York: North Holland, 1980.

Shannon, R.E. Systems simulation: The art and science. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Sheldon, M.S., & Zagorski, H.J. Project NORM Mission I Research Report (TM-2232/001/00). Santa Monica, Calif.: System Development Corp., October 1965.

Siegel, A.I., & Federman, P.J. Mental load analysis for man-machine systems (Technical Report). Wayne, Pa.: Applied Psychological Services, Inc., 1980.

Siegel, A.I., & Wolf, J.J. Man-machine simulation models. New York: Wiley, 1969.

Swain, A.D., & Guttman, H.E. Handbook of human reliability analysis with emphasis on nuclear power plant applications (NUREG/CR-1278). Albuquerque, N.M.: Sandia Laboratories, October 1980.

Turing, A.M. Computing machinery and intelligence. Mind, 1959, 59, 103-127.

Zanakis, S.H. Extended pattern search with transformations for the three-parameter Weibull MLE problem. Management Science, 1979, 11, 1149-1161.

APPENDIX

Clarification of Operator Following Response Qualifications

Subscript	Explanation
A_1	Allowable if engagement requires hooking a different track from that apparently hooked, or being hooked.
A_2	Allowable if Alert Process transition requires preempting current state.
A_3	Allowable if blinking track number in TBEQ reprioritizes operator state-transition requirements.
A_4	Allowable if TBEQ process requirements preempts operator state-transition requirements.
A_5	Allowable under special conditions [to be determined (TBD)]
A_6	Allowable if conditions warrant joystick usage.
A_7	Allowable if track requiring operator action is not in the first line of the pre-engagement portion of the TBEQ.
A_8	Allowable if operator requires a situation configuration change.
A_9	Allowable if and only if a track has previously been hooked.
E_1	Error unless a hooked track's priority (that is in the TBEQ) is being reviewed prior to engagement.